A single-cell multi-omic and spatial atlas of nodal B-cell lymphomas reveals B-cell maturation drives intratumor heterogeneity

Donnacha Fitzgerald [1 2 3], Tobias Roider [1 2], Marc-Andrea Baertsch [2 4], Artur Kibler [6 7], Anastasiia Horlova [3], Erin Chung [3], Harald Vöhringer [1 2 3], Anna Mathioudaki [1 2 8], Bettina Budeus [6], Verena Passerini [9], Mareike Knoll [2], Johannes Mammen [1 2], Linsha Li [3], Léandra Caillé [2], Felix Czernilofski [1 2], Peter-Martin Bruch [1 2 5 10], Nora Liebers [1 2 5 10 11], Matthias Meyer-Bender [1 3], Oliver Weigert [9 11], Judith Zaugg [1 8], Garry Nolan [4], Marc Seifert [5 6 7], Wolfgang Huber [1 3 12*], Sascha Dietrich [1 2 5 10 12*]

[1] European Molecular Biology Laboratory (EMBL), Molecular Medicine Partnership Unit (MMPU), Heidelberg, Germany
[2] Department of Hematology, Oncology and Rheumatology, Heidelberg University Hospital, Heidelberg, Germany
[3] European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany
[4] School of Medicine, Department of Microbiology and Immunology, Stanford University, Stanford, CA, USA
[5] Department of Oncology, Hematology and Clinical Immunology, Medical Faculty of Heinrich-Heine-Universität, Düsseldorf University Hospital, Düsseldorf, Germany
[6] Institute of Cell Biology, University of Duisburg-Essen, Essen, Germany
[7] Department of Molecular Genetics, Essen University Hospital, Essen, Germany
[8] European Molecular Biology Laboratory (EMBL), Structural and Computational Biology Unit, Heidelberg, Germany
[9] Medical Department III, Ludwig-Maximilians-University (LMU) Hospital, Munich, Germany
[10] Center for Integrated Oncology Aachen-Bonn-Cologne-Düsseldorf (CIO ABCD), Aachen Bonn Cologne Düsseldorf, Germany
[11] Department of Translational Medical Oncology, National Center for Tumor Diseases (NCT) Heidelberg and German Cancer Research Center (DKFZ), Heidelberg, Germany
[12] These senior authors contributed equally to this study

*Correspondence: sascha.dietrich@embl.de, wolfgang.huber@embl.org

Abstract

Intratumor heterogeneity is intrinsic to cancer pathogenesis and evolution, although little is known about how it relates to the differentiation trajectories of the tumor's cell-of-origin. Nodal B-cell non-Hodgkin lymphomas are a diverse set of cancers thought to originate from distinct stages of B-cell maturation. Through a single-cell multi-omic and spatial atlas of diffuse large B-cell, mantle cell, follicular, and marginal zone lymphomas along with reactive lymph nodes (n=51), we found multiple B-cell maturation states coexist within the same tumors. Intratumor maturation states emerged from the same cell-of-origin, revealing that maturation remains plastic in malignancy. The maturation state composition varied across entities and tumors, which included mixtures of cell-of-origin subtypes. Intratumor maturation states inhabited unique spatial niches, which typically retained their maturation-associated cellular interactions and regulatory networks. Intratumor maturation states showed distinct expression patterns of genetic variants, suggesting that maturation and genetic aberrations are intertwined. Our findings put forward a transformative model for cancer pathogenesis, where differentiation continues in malignancy and is central to tumor heterogeneity and evolution.

Introduction

Intratumor heterogeneity poses significant implications for prognosis and treatment response across cancers[1]. Intratumor heterogeneity can manifest at cellular, genetic, epigenetic, and microenvironmental levels, contributing to complex tumor architecture and adaptive resistance mechanisms[2,3]. Importantly, greater intratumor heterogeneity has been associated with a worse prognosis, due to the increased likelihood of therapy-resistant subclones[4–6]. Thus, a better understanding of what drives intratumor heterogeneity promises to improve our understanding of cancer pathogenesis, evolution, and resistance mechanisms.

There is increasing evidence that the cell-of-origin, the primary cell type acquiring the first genetic aberrations leading to cancer, influences tumor behavior and response to therapy[7]. The molecular and functional features that are inherited by the tumor cells from their normal counterpart have clinical relevance for many cancer types[8]. Nodal B-cell non-Hodgkin lymphomas (B-NHL), a heterogeneous set of malignancies causing over 200,000 deaths globally each year[9], have been well studied in this regard. B-cell-mediated immunity is honed in secondary lymphoid organs, such as lymph nodes, that are focal points in the B-cell maturation process. Upon T-cell-dependent activation, B-cells migrate into the B-cell follicles to initiate the germinal center (GC) reaction. Here, B-cells proliferate and undergo somatic hypermutation as dark zone centroblasts (DZ), and are then selected for improved affinity of the B-cell receptor (BCR) as light zone centrocytes (LZ) by antigen-presenting cells (APCs) such as T-follicular helper (TFH) cells and follicular dendritic cells (FDC). Throughout this iterative process, GC B-cells may be stimulated by APCs to differentiate into memory B-cells (Mem) or plasma cells[10,11]. B-NHL entities include typically GC-origin diffuse large B-cell lymphoma (DLBCL, GCB) and follicular lymphoma (FL), post-GC or activated B-cell DLBCL (DLBCL, non-GCB/ABC), as well as mixed-origin mantle cell lymphoma (MCL) and marginal zone lymphoma (MZL). They are thought to originate from B-cells at different stages of the maturation process that predominantly reside in these specialized environments, such as DZ or LZ states in GC-origin DLBCL and FL, and Mem (or Naïve/Plasma) states in non-GCB DLBCL, MCL, and MZL[11–13].

These B-NHL entities have variable growth rates and clinical courses, ranging from the more indolent FL and MZL to the more aggressive MCL and DLBCL[14]. In addition, there are considerable biological and clinical variabilities between patients among patients with the

same B-NHL entity[15,16]. This intertumor heterogeneity is linked to the maturation state of the tumor's normal counterpart. For example, activated DLBCL (ABC-type) from post-GC B-cells and germinal center DLBCL (GCB-type) from GC B-cells are two major DLBCL subtypes associated with different maturation states[15]. Following standard treatment with chemoimmunotherapy, ABC-DLBCL shows more frequent relapse compared to GCB-DLBCL. This association between DLBCL tumor maturation state and clinical outcomes has recently been extended beyond the ABC vs GCB dichotomy to across the maturation spectrum[17]. In parallel, intratumor subpopulations show differences in their treatment sensitivities[3,18,19].

While it has been well established that B-cell maturation significantly contributes to inter-tumor heterogeneity across B-NHL entities, we hypothesize that maturation also drives intratumor heterogeneity. This would shed new light on the role of differentiation trajectories in cancer pathogenesis, variation, and evolution.

Here, we test this hypothesis in B-NHL. We first construct a single-cell reference map of B-cell maturation states in reactive lymph nodes (rLN) using flow cytometry, RNA-sequencing, and cellular indexing of transcriptomes and epitopes (CITE-Seq)[20]. We then leverage this to classify the B-cell maturation state of tumor cells and profile their distribution across tumors and entities. We characterize the regulatory networks distinguishing maturation states in malignancy and genetic variation between these states. As the follicular architecture of the lymph node mediates the B-cell maturation process, we use multiplexed tissue imaging with DNA-conjugated antibodies (CODEX)[21] to study the spatial context of tumor maturation states. Through a comprehensive single-cell exploration of B-NHL, spanning transcriptional, proteomic, epigenetic, genetic, and spatial facets, we bring to light a novel paradigm of tumor evolution, whereby the tumor's differentiation state remains plastic and a major axis of intratumor heterogeneity.

Results

Sample overview

We obtained 51 lymph node (LN) samples from patients diagnosed with mantle cell lymphoma (MCL, n = 8), follicular lymphoma (FL, n = 12), germinal center (GCB, n = 5) or activated B-cell (non-GCB/ABC, n = 7) diffuse large B-cell lymphoma (DLBCL), marginal zone lymphoma (MZL, n = 11), or non-malignant reactive lymph nodes (rLN, n = 8). Of the malignant LN samples, 20 were collected at the time of initial diagnosis and 23 were from patients who had previously undergone one or more lines of systemic treatment. Relapse samples were collected at least 3 months after cessation of systemic treatment. We summarized patient characteristics in Supplementary Table 1.

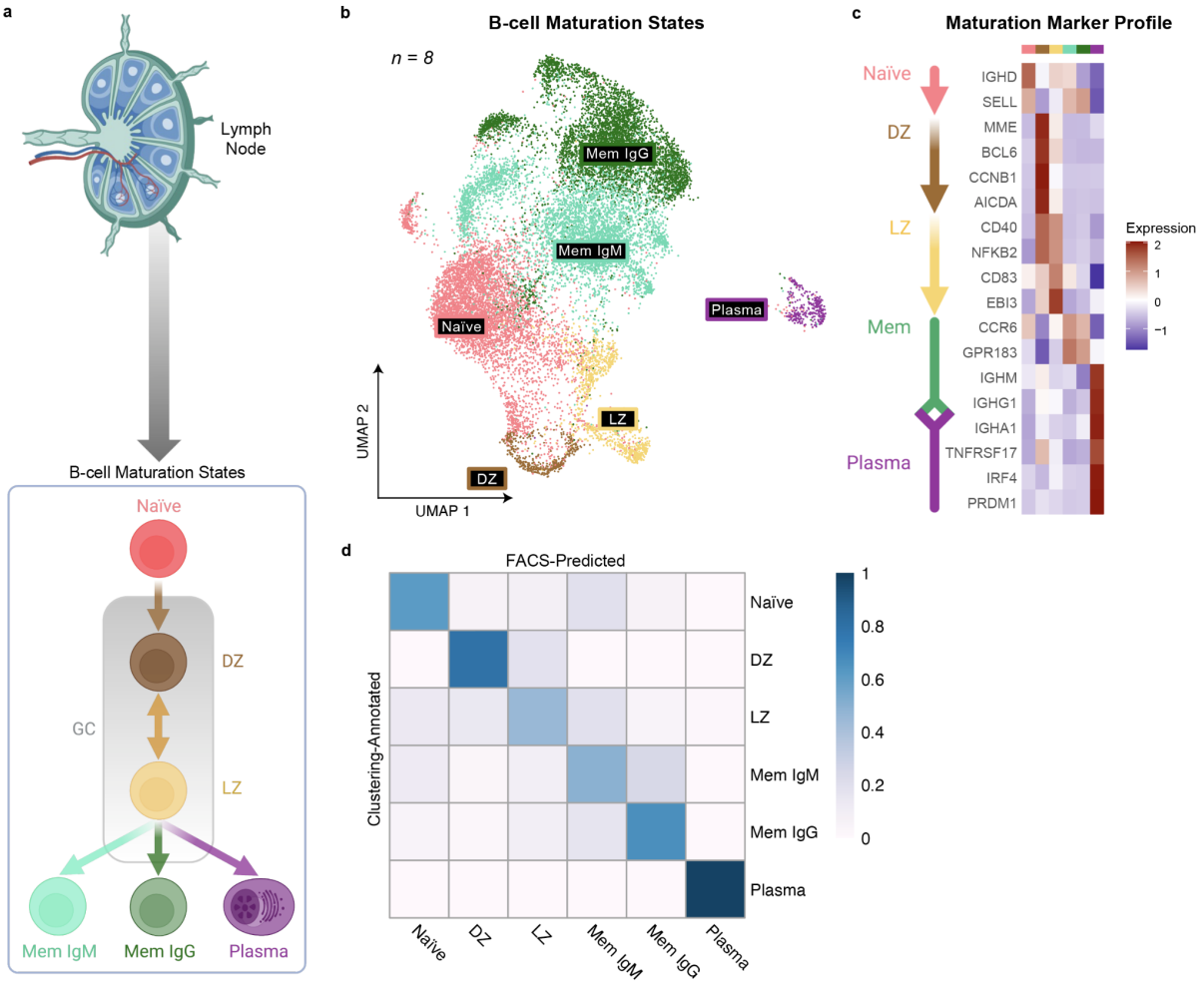
We characterized these samples with CITE-Seq[20], a multi-modal method coupling single-cell RNA sequencing with surface protein profiling with oligonucleotide-tagged antibodies (n = 70, Supplementary Table 2). We aligned transcripts to the hg38 reference genome. After quality control, we obtained data for 154,282 B-cells with a median of 2,988 B-cells per sample [140, 7,868] and a median of 6,887 transcript and 2,532 surface protein counts per cell.

A single-cell B-cell maturation reference map in reactive lymph nodes

To characterize B-cell maturation states in the CITE-Seq data, we first sorted B-cells from five rLN samples into different maturation state subsets using fluorescence-activated cell sorting (FACS). The gating scheme is specified in Extended Data Fig. 1. We identified naïve, germinal dark zone centroblasts (DZ), light zone centrocytes (LZ), IgD+/IgM+ memory (Mem IgM), class-switched memory of predominantly IgG class (Mem IgG), and plasma maturation states. Each of these subsets was present in each sample. We then characterized these six maturation states from each sample with bulk RNA sequencing (Extended Data Fig. 2a-c). We trained a logistic regression classifier on the resulting gene expression profiles to predict maturation states (balanced accuracy in cross-validation = 94%), which we then applied to the CITE-Seq data from all 8 rLN samples (16,625 cells) using the single-cell transcriptomic profiles (Extended Data Fig. 2e-d).

In parallel, we applied Louvain clustering to the transcriptomic data in this rLN dataset. We obtained 16 clusters and annotated these to the six B-cell maturation states based on the presence of established markers[10,11,17,22–26] (Supplementary Table 3, Fig. 1b-c). For each cluster, the annotated maturation state corresponded with the most prevalent prediction for the cluster's constituent cells from the logistic regression classifier (Fig. 1d, Extended Data Fig. 2e). We observed that the rLN samples were composed of 60% memory (split equally between IgD+/IgM+ (Mem IgM), and class-switched (Mem IgG)), 30% naïve, 6% light zone centrocytes (LZ), 3% dark zone centroblasts (DZ), and 2% plasma cells.

Fig. 1: A single-cell B-cell maturation reference map in reactive lymph nodes



a, Schematic of the B-cell maturation trajectory in the lymph node. The labeled populations represent the B-cell maturation states characterized by FACS and CITE-Seq. Illustrations were created with BioRender.com[27]. **b**, Transcriptomic UMAP of the rLN reference CITE-Seq dataset (8 samples) labeled by the B-cell maturation states in a. Transcriptomic clusters were assigned to maturation states based on their expression of the maturation markers in Supplementary Table 3. **c**, Heatmap showing the z-scored average expression of a subset of markers for each maturation state annotated in the reference CITE-Seq dataset. **d**, Confusion matrix of cells'

maturation state labels annotated by maturation marker profiling of transcriptomic clusters (y-axis) and predicted by a logistic regression[28] classifier trained on RNA-sequencing data from B-cell maturation states sorted with FACS(x-axis)(Extended Data Fig. 1-2). The color scale shows probability estimates for each class. Maturation state annotations: Naïve = Naïve B-cells, DZ = Centroblasts from the dark zone of the germinal center, LZ = Centrocytes from the light zone of the germinal center, Mem IgM = IgD+ and IgM+ memory B-cells, Mem IgG = class-switched (IgG+ or IgA+) memory B-cells, Plasma = plasma cells.

Divergence of B-cell maturation in tumors

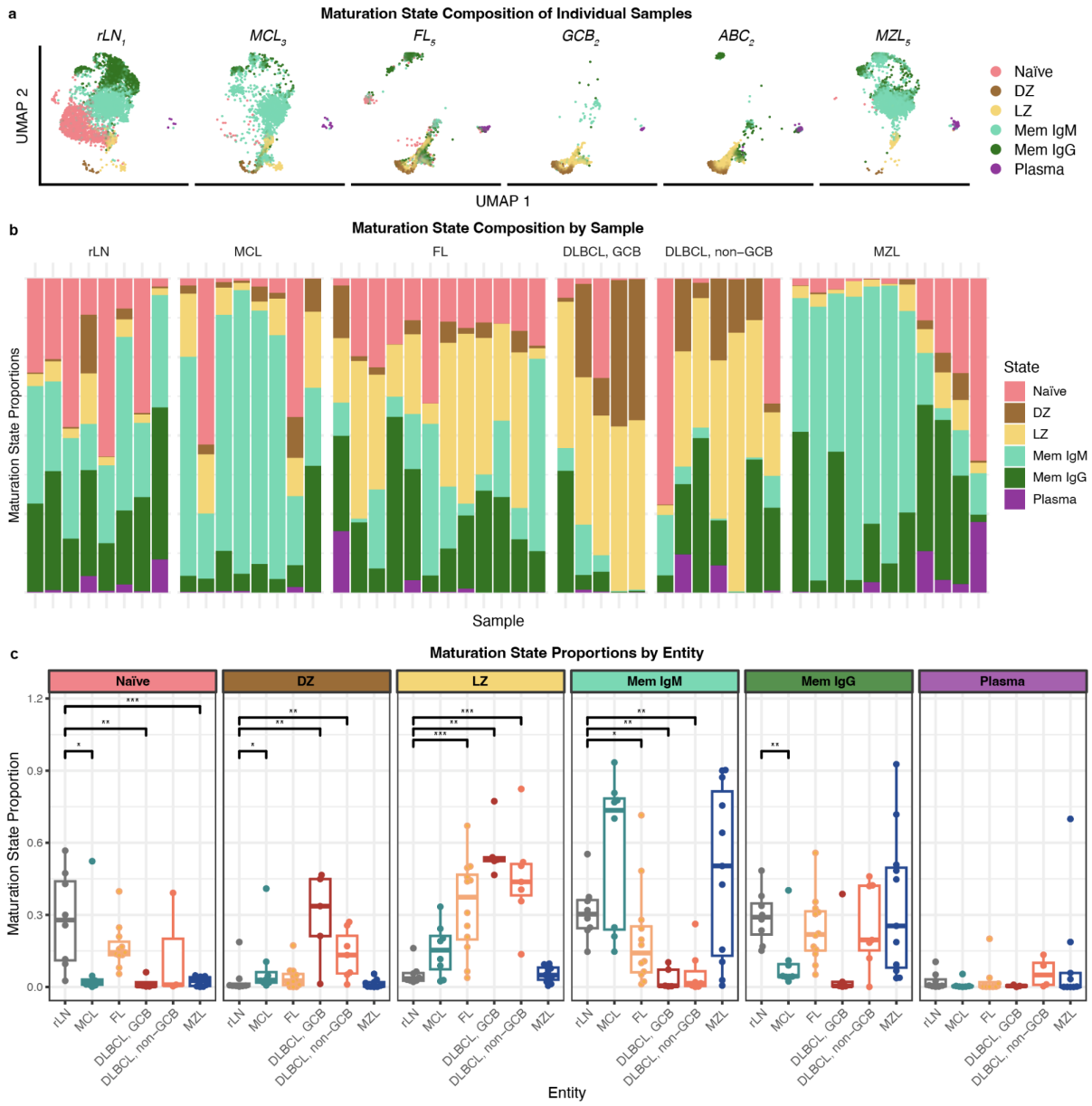
Having characterized B-cell maturation states in non-malignant lymph nodes, we then sought to profile these states in tumors using the 43 B-NHL samples in our CITE-Seq cohort (8 MCL, 12 FL, 5 GCB DLBCL, 7 non-GCB DLBCL, and 11 MZL). For each sample, we identified the malignant cells based on light chain restriction[29], whereby we considered transcriptional clusters with a kappa or lambda light chain proportion >0.75 as malignant. We verified this approach by applying B-cell receptor (BCR) profiling to 8 of the samples. We found that each tumor sample contained a single expanded B-cell receptor clone, representing the malignant cells, with a restricted immunoglobulin light chain. Non-malignant B-cells, comprising mostly naïve B-cells, represented a median of 6% [0%, 94%] of all B-cells across tumor samples (Extended Data Fig. 3).

To classify B-cell maturation states in the tumor samples, we leveraged a data integration approach based on mutual nearest neighbors and canonical correlation analysis[30] to map maturation states from the rLN reference dataset in each tumor. The maturation states classified in malignant cells corresponded with their maturation marker profiles. Furthermore, using the 50 most differentially expressed genes for each maturation state in a published tonsil dataset[17], we calculated maturation state gene signature scores in our B-NHL dataset. As with the maturation marker profiles, gene signature scores of maturation states in malignant entities reflected scores in their respective states in the rLN dataset (Extended Data Fig. 4a-b).

For subsequent analyses, we focused only on the malignant B-cells in tumor samples. We expected to find a single B-cell maturation state per tumor, reflecting a fixed maturation state inherited from the cell-of-origin. Instead, we detected a spectrum of maturation states within individual tumors (Fig. 2a-b, Extended Data Fig. 5). As intratumor maturation states were of

a single BCR clone, they diverged from a common cell-of-origin. These observations indicate that B-cell maturation is not fixed in malignancy, but remains plastic and divergent.

Fig. 2: Tumors consist of multiple B-cell maturation states



a, Transcriptomic UMAPs for individual samples from each entity, labeled by B-cell maturation states assigned by label transfer from the reactive lymph node reference (Fig. 1b). Only malignant cells are shown for tumor samples. **b**, Maturation state composition of all samples (n=51) split by entity and ordered by days since diagnosis. **c**, Box plot showing the proportion of each maturation state in each entity. Each data point is a sample, grouped by entity. The Wilcoxon signed-rank test[31] was performed for each maturation state's proportion between reactive lymph nodes and each entity: $p<0.05$ (*), $p<0.01$ (**), $p<0.001$ (***). rLN = reactive lymph nodes, MCL = mantle cell lymphoma, FL = follicular lymphoma, DLBCL = diffuse large B-cell

lymphoma (GCB = germinal center, non-GCB/ABC = activated B-cell), MZL = marginal zone lymphoma. See Fig. 1 for maturation state annotations.

Maturation states are a source of intratumor and intertumor heterogeneity

We observed a characteristic spectrum of maturation states in each entity; the predominant states reflected their associated cell-of-origin, such as GC states in DLBCL GCB and memory states in MCL and MZL. However, we observed substantial variation in maturation state proportions between samples of the same entity (Fig. 2b). Predicting the samples' entity by maturation state composition achieved a maximum accuracy of 63% (logistic regression, nested cross-validation) (Extended Data Fig. 4c). This highlights B-cell maturation as a major contributor to both intratumor and intertumor variation.

FL and non-GCB DLBCL tumors showed diverse mixtures of GC (DZ and LZ) and post-GC (Mem and Plasma) states, suggesting that, like DLBCL, FL may also be capable of transformation into post-GC phenotypes. Both DLBCL and FL showed significant enrichment in the LZ state compared to rLN controls ($p < 0.01$, Wilcoxon signed rank test), but FL tumors were not enriched for the DZ state (Fig. 2b), implying that DLBCL better maintains a DZ phenotype. Continuation of somatic hypermutation and proliferation in the DZ state may promote DLBCL's characteristically more aggressive disease course[32].

To study how the phenotypic diversity in B-NHL tumor cells could be explained by variation in B-cell maturation states, we took an unsupervised approach to multimodal subpopulation mapping on the full CITE-Seq B-cell dataset (51 samples, 154,282 cells). We created an integrated dimensionality reduction and clustering from all B-cells' transcriptomic and surface protein features using latent factors from Multi-Omic Factor Analysis (MOFA)[33] (Extended Data Fig. 6a). The 25 multimodal subpopulations identified were segregated predominantly by the maturation state, supporting maturation as a major driver of tumor variation (Extended Data Fig. 6b-e). In addition, multiple distinct subpopulations were observed within each maturation state. The greatest heterogeneity was observed among memory B-cell states, in which subpopulations were segregated by B-NHL entity. Distinguishing features among these subpopulations pointed towards known pathogenic mechanisms, such as *CCND1* overexpression in MCL (Extended Data Fig. 6f-g).

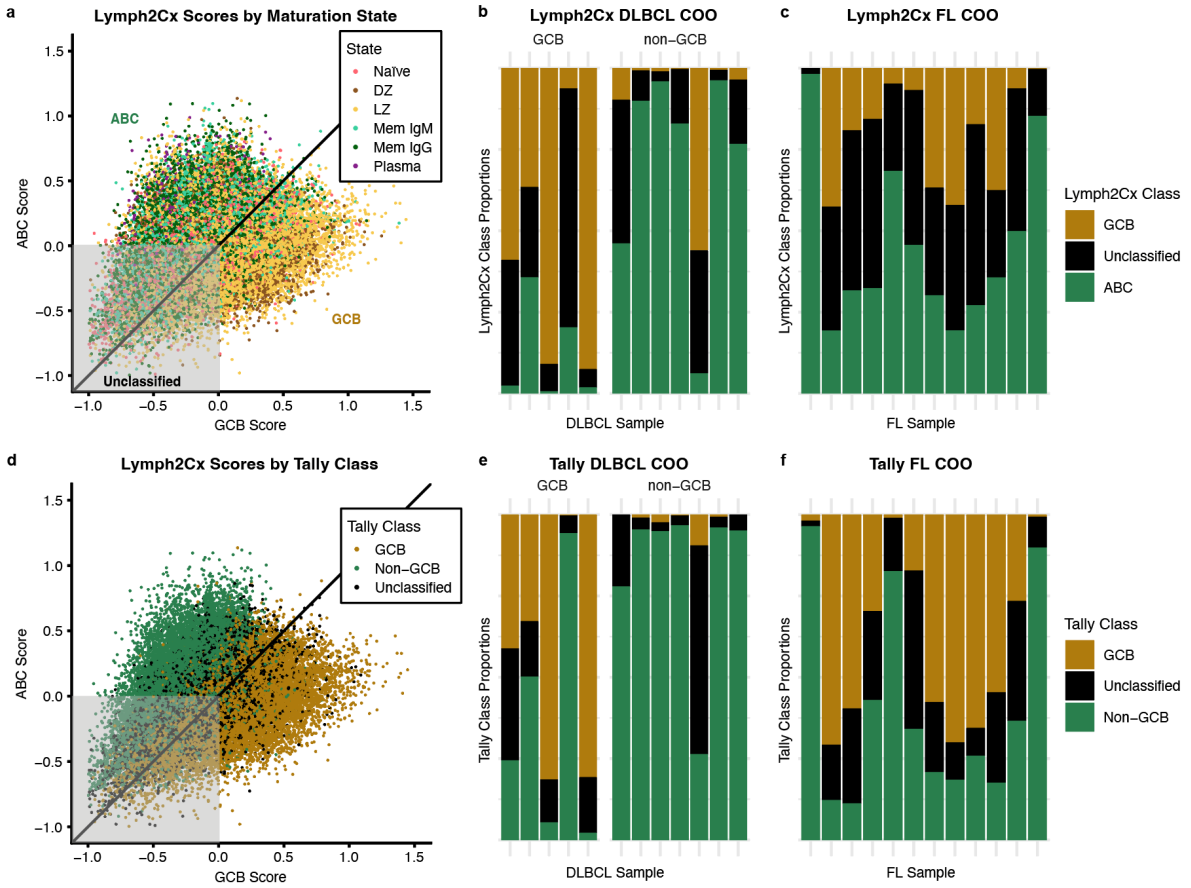
Cell-of-origin classification reveals multiple subtypes within each tumor

Next, we examined the implications of intratumor variation in maturation states for established diagnostic cell-of-origin (COO) classifiers. The Lymph2Cx gene expression classifier[34] is used to classify DLBCL into germinal center (GCB) and activated B-cell (ABC) subtypes based on a 20-gene panel associated with GC and post-GC maturation states. We applied this classifier to the single-cell RNA-sequencing (scRNA-seq) data from the malignant cells of each DLBCL tumor, where higher GCB and ABC gene expression signature scores corresponded with GC (DZ and LZ) and post-GC (Mem and Plasma) states, respectively (Fig. 3a). 20-30% of tumor cells did not display a dominant GCB or ABC score and remained unclassified. Rather than a single GCB or ABC class per tumor, several tumors consisted of multiple cell-of-origin subtypes (GCB, unclassified, and ABC), whereby the dominant class largely reflected the GCB or non-GCB diagnosis. However, the presence of multiple classes on the single-cell level indicates the pathology of DLBCL is more complicated than previously thought, with a mixture rather than a single cell-of-origin subtype for each tumor (Fig. 3b).

We confirmed this observation with the Tally DLBCL COO classifier, a simpler classifier typically employed via immunohistochemistry analysis using the presence or absence of proteins associated with GCB (CD10, GCET1, and LMO2) or non-GCB (MUM1 and FOXP1) states to classify tumors into GCB or non-GCB subtypes[35]. This revealed a similar pattern to that of applying the Lymph2Cx classifier to our scRNA-seq data. With both classifiers, GCB DLBCL often contained populations of ABC-like cells, consistent with the B-cell maturation process transforming a portion of tumor cells of GCB origin to an ABC subtype (Fig. 3d-e).

FL showed both GC and post-GC states, although FL is not known to include an ABC subtype like DLBCL. Thus, we examined whether both GCB and ABC subtypes could also be detected at the single-cell level in FL tumors. The application of the Lymph2Cx and Tally classifiers revealed the presence of both GCB and ABC classes in 10 out of 12 FL tumors (Fig. 3c+f), revealing the emergence of an ABC-like subtype in FL tumors.

Fig. 3: Cell-of-origin classification reveals multiple subtypes within each tumor



a, Normalized GCB and ABC scores for each DLBCL and FL tumor cell determined by the Lymph2Cx cell-of-origin classifier, labeled by B-cell maturation state. The diagonal line divides cells by their classification (above = ABC, below = GCB). Cells without ABC or GCB gene set average expression above the housekeeping genes were unclassified (grey box). **b-c**, Lymph2Cx class proportions among the tumor cells of DLBCL samples faceted by GCB or non-GCB diagnosis (b) and FL samples (c). **d**, Normalized GCB and ABC scores for each DLBCL and FL tumor cell determined by the Lymph2Cx classifier, as shown in (a), labeled by Tally class. **e-f**, Tally class proportions among the tumor cells of DLBCL samples faceted by GCB or non-GCB diagnosis (e) and FL samples (f). FL = follicular lymphoma, DLBCL = diffuse large B-cell lymphoma (GCB = germinal center, non-GCB = non-germinal center), COO = cell-of-origin. See Fig.1 for maturation state annotations.

Longitudinal patterns in tumor maturation state composition

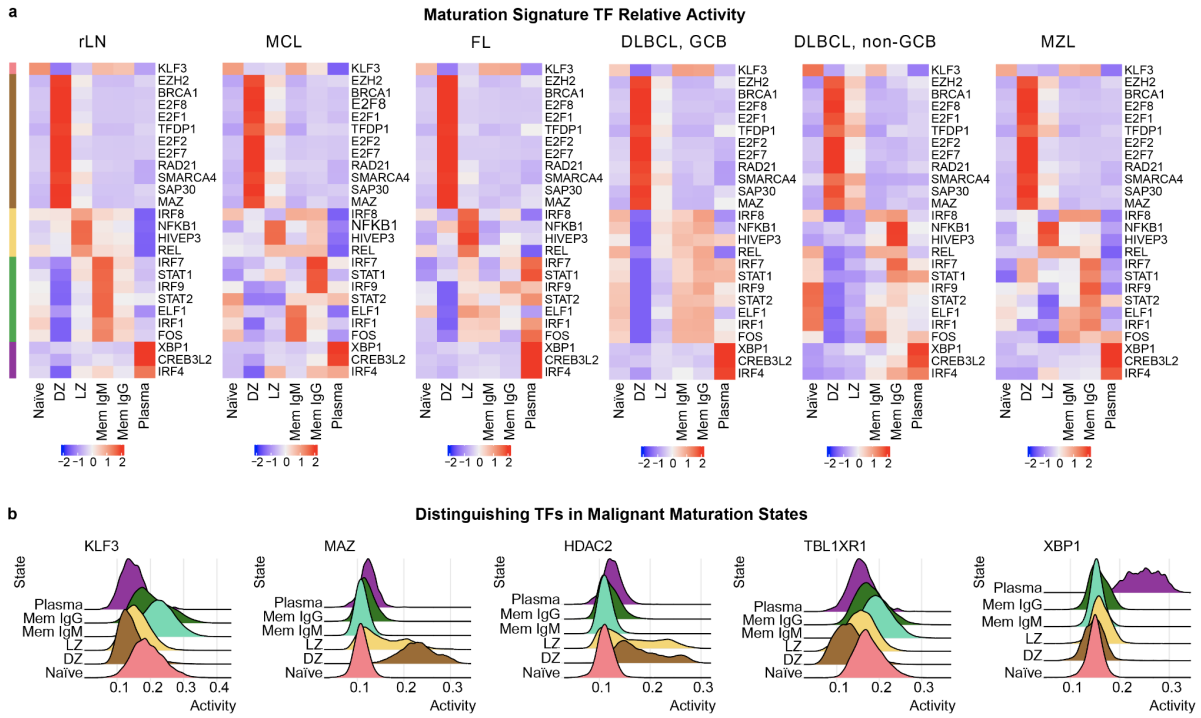
Based on these findings, we hypothesized that intratumor heterogeneity introduced by B-cell maturation may play a role in tumor evolution, whereby a shift in maturation state composition may occur over time due to progression in maturation or selection by treatment. This was supported by our observed increases in the class-switched memory (Mem IgG) and plasma states in MZL samples and an increase in the ABC cell-of-origin subtype in FL samples taken later from diagnosis (Fig. 2b). We explored this in longitudinal lymph node

samples from 3 B-NHL patients (Supplementary Table 1). In an MZL tumor from a patient who had a complete response to 6 cycles of obinutuzumab and bendamustine but relapsed 15 months later, we observed a decrease in plasma cells with an increase in the Mem IgG state (Extended Data Fig. 7a). We also observed enrichment in plasma cells after 11 months in a GCB DLBCL tumor from a patient who relapsed following CAR-T cell therapy (axicabtagene ciloleucel) (Extended Data Fig. 7b). These results point to the possible progression of the maturation process over time. However, this effect was probably largely time-dependent, as we observed little change in maturation state composition after 2 months in another DLBCL patient who relapsed to the same CAR-T cell therapy (Extended Data Fig. 7c).

Transcription factor signatures of maturation states maintain differential activity in malignancy

Transcription factors (TFs) are key mediators of the B-cell maturation process, although alteration of the epigenomic landscape has been described in the pathogenesis of B-NHL[12,22,36–38]. We investigated whether differences in TF activities between maturation states were maintained in malignancy, which could enable a tumor to diverge into multiple maturation states despite epigenomic dysregulation. We inferred TF activities computationally from the scRNA-seq data (51 samples) using the *SCENIC* workflow[39]. Here, we identify coexpression modules between TFs and candidate target genes with an enriched binding motif and score TF activity based on the modules' expression in each cell. We defined signature maturation TFs as those with differentially expressed target genes between maturation states in the rLN reference dataset (BH-adjusted p-value < 10e-16, average log fold-change > 0.4). We calculated the relative activity of these TFs between states for each entity. TFs that were differentially active between maturation states in rLN also showed similar activity patterns across states in malignant entities (Fig. 4a), suggesting that TF maturation state signatures were predominantly conserved across malignancies. This pattern was partially lost, however, in the LZ state in DLBCL (both GCB and non-GCB), where TFs elevated in the rLN LZ (*IRF8, NFKB1, HIVEP3,* and *REL*) showed higher activity in the memory state. Large variation was observed in the relative activity of memory signature TFs across entities (eg. *IRF7, STAT1, IRF9*). However, this corresponds with the phenotypic diversity of nodal memory B-cells, which have already completed the maturation process and are therefore less likely to be dependent on maturation GRNs.

Fig. 4: Maturation signature TFs maintain differential activity in malignancy



a, Scaled activity of transcription factors (TFs) across B-cell maturation states in each entity inferred using the *scenic* python package[39]; only TFs significantly enriched (log2 fold-change >0.4, p < 10e-16 as determined with the *MAST* R package[40]) for B cell maturation stages target genes in reactive lymph node samples are shown. TFs (y-axis) are ordered by the maturation state in reactive lymph nodes in which they are enriched. Non-malignant cells in tumor samples were excluded based on light-chain restriction. **b**, Density plots comparing the activity distribution of a subset of distinguishing transcription factors in each tumor maturation state in malignant cells aggregated from all tumor samples (n = 43). rLN = reactive lymph node, MCL = mantle cell lymphoma, FL = follicular lymphoma, DLBCL = diffuse large B-cell lymphoma (GCB = germinal center, non-GCB = non-germinal center), MZL = marginal zone lymphoma. See Fig. 1 for maturation state annotations.

TFs with the greatest differences in activity between maturation states in malignancy were primarily those involved in B-cell development, activation, and differentiation. These included *KLF3* in naïve and memory B-cells, *MAZ* and *HDAC2* in GC B-cells, *TBL1XR1* in memory, and *XBP1* in plasma cells (Fig. 4b and Extended Data Fig. 8). These may play a central role in driving or maintaining the diversity of intratumor maturation states.

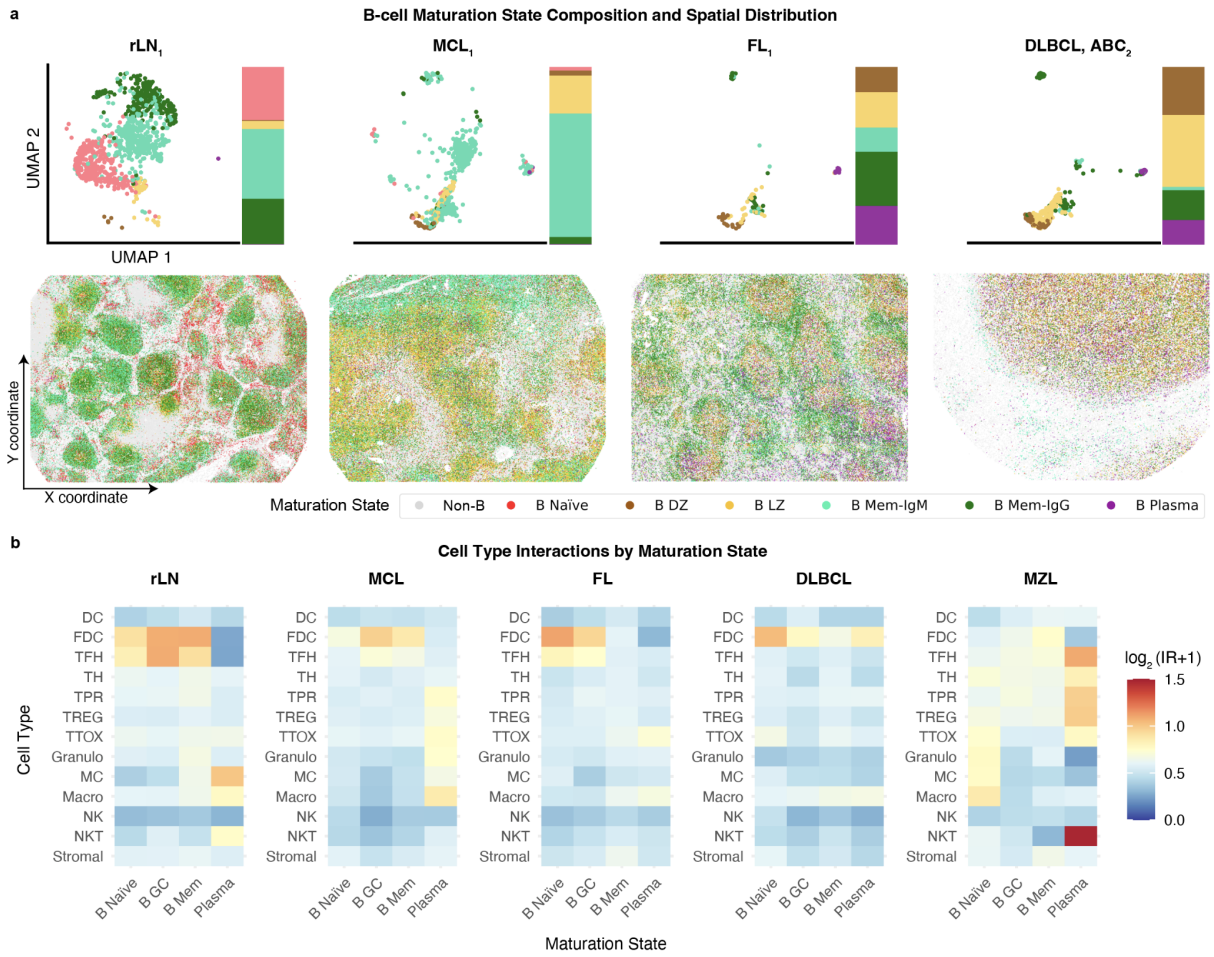
Intratumor maturation states occupy distinct spatial microenvironments

The follicular architecture of lymph nodes, which is central to the B-cell maturation process, is often disrupted in malignancy. Therefore, we sought to understand the spatial context of

intratumor maturation states and whether spatial niches could still facilitate maturation in malignancy. We characterized the spatial distribution of cell types with CODEX[21] (52 features, Supplementary Table 2) for 19 samples (29 slides) in the CITE-Seq cohort. Microenvironmental cell types were defined according to previously published nodal cell type marker profiles[41]. Using logistic regression, we transferred B-cell maturation state labels in the CITE-Seq dataset samplewise to B-cells in the CODEX dataset via the shared protein features (n=28) between both datasets (Supplementary Table 2). Although we employed random sampling for class balancing, we obtained a high correlation in maturation state proportions between the CITE-Seq and CODEX datasets (median R=0.91, p = 0.011) (Extended Data Fig. 9).

We observed substantial variation in tumor microenvironment composition and structure among tumors and entities, typically with a loss of the normal follicular structure in MCL and DLBCL (Extended Data Fig. 10a). Tumor cells of different maturation states still tended to segregate spatially (Extended Data Fig. 10b). For example, FL_1 consisted of expanded follicles with DZ and LZ states in the GC and surrounded by memory and plasma states. MCL_1 consisted of spatially distinct LZ and memory states, albeit without a typical follicular structure. ABC2 showed a spread of several maturation states with little spatial distinction, consistent with DLBCL's diffuse nature (Fig. 5a). These variations highlight the variable degrees and patterns of spatial compartmentalization between entities.

Fig. 5: Intratumor maturation states occupy distinct spatial microenvironments



a, UMAP of scRNA-seq data labeled by B-cell maturation states mapped from the reactive lymph node reference with B-cell maturation state proportions (top), and spatial distribution of B-cell maturation states in a reactive lymph node (rLN), follicular (FL), mantle cell (MCL), and non-germinal center diffuse large B-cell lymphoma (DLBCL, non-GCB) sample from CODEX[21] images (52 markers) on FFPE tissue sections (bottom). See Extended Data Fig. 10 for the distribution of maturation states and cell types on all CODEX slides (n=29). **b**, Log_2 of the pairwise cell-cell observed over expected interaction ratios (IR) between B-cell maturation states and all other cell types in each entity, with a pseudocount of 1. Spatial interactions were determined with Delaunay triangulation. A higher ratio is associated with increased proximity. B Naïve = naïve B-cells, B DZ = centroblasts from the dark zone of the germinal center, B LZ = centrocytes from the light zone of the germinal center, B Mem IgM = IgD+ and IgM+ memory B-cells, B Mem IgG = class-switched (IgG+ or IgA+) memory B-cells, B Plasma = plasma cells, CD4T_naive = naive CD4+ T-cells, CD8T_naive = naive CD8+ T-cells, TH_memory = memory helper T-cells, TTOX_memory = memory cytotoxic T-cells, TTOX_exh = exhausted cytotoxic T-cells, NKT = natural killer T-cells, TFH = follicular helper T-cells, TPR = proliferating T-cells, TREG = regulatory T-cells, FDC = follicular dendritic cells, DC = dendritic cells, Macro = macrophages, Stromal = stromal cells, NK = natural killer cells, MC = monocytes, Granulo = granulocytes.

We explored spatial niches in the CODEX data with K nearest neighbor (KNN) graph-based cellular neighborhood (CN) analysis[42]. We identified 11 CNs using 20 KNNs and characterized them based on the cell types enriched in each CN (Extended Data Fig. 11). We observed that CNs common in rLNs were also present in tumor samples, including T-cell zones (Th Mem zone and Mixed T zone) and a stromal zone bordering B-cells in tumors. The majority of tumors' malignant B-cells, however, were found in CNs different from the follicular CN found in rLN samples. Each tumor maturation state was dominant in a different CN, which often co-existed in the same sample. For example, FL_1 consisted of DZ, LZ, and plasma-enriched zones. We observed that intratumor maturation states had different cellular microenvironments with unique immune compositions, such as enrichment of exhausted cytotoxic T-cells (TTOX_exh) and macrophages (Macro) in plasma tumor zones (Extended Data Fig. 11). Tumor infiltration of immune cells such as cytotoxic T-cells is central to the success of cellular immunotherapies. Compared to other maturation states, plasma tumor cells were more frequently neighboring memory cytotoxic T-cells (TTOX_memory).

Heterogeneous microenvironments may promote divergence of maturation states by enabling or blocking different cellular interactions crucial to stages of the maturation process. FDCs and TFH were enriched in GC (DZ and LZ) predominant tumor CNs, potentially explaining how the differentiation of malignant GC cells into post-GC memory or plasma cells occurred (Extended Data Fig. 11b). To quantify the extent to which cell-cell interactions were maintained in each entity, we computed the pairwise cell-cell observed over expected interaction ratios (IR) between B-cell maturation states and other cell types in each entity[42]. We found that GC B-cell interactions with FDC and TFH were partially maintained in tumors, therefore enabling the GC reaction to continue in malignancy (Fig. 5b). Although this was to a lesser extent in the more diffuse DLBCL and generally post-GC MZL, which could explain their lower diversity of maturation states compared to FL (Fig. 2b).

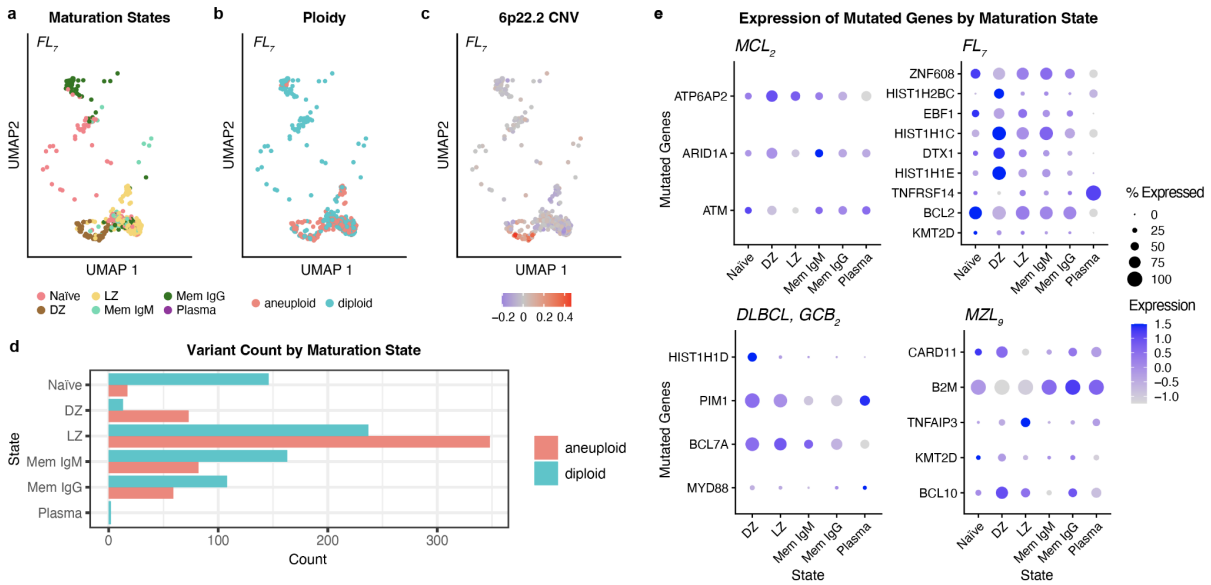
Genetic variants among intratumor maturation states

Several characteristic features of B-cell maturation states are known to be commonly aberrated in B-NHL (eg. *BCL6, IRF4,* and *PRDM1*)[10,43]. We considered whether different intratumor maturation states may be associated with subclonal genetic variation. We profiled genetic variants, including single-nucleotide variants (SNVs), insertions and deletions, and copy number variants (CNVs) from targeted DNA-sequencing data in bulk tumor samples.

We then inferred CNVs on the single-cell level from gene expression data using the *copykat* method[44].

We observed distinct CNVs among cells in individual tumors, showing that a tumor's CNVs could be confined to a portion of the tumor cells. Some CNVs arose between different B-cell maturation states. For example, in FL_7, most GC tumor cells showed aneuploidy compared to the predominantly diploid non-GC tumor cells (Fig. 6a-d). This tumor showed copy number gain in the 6p22.2 region (Supplementary DNA Sequencing Report), which is associated with the disease progression of FL[45]. 6p22.2 copy number gain was restricted to the DZ tumor cells (Fig. 6c), implying that 6p22 gain may either lock cells in the DZ stage of maturation, or DZ cells may be more likely to acquire this copy number variant.

Fig. 6: Genetic variants among intratumor maturation states



a-c, Reference-based UMAP of malignant cells from FL_7 labeled by (a) B-cell maturation state, (b) diploid or aneuploid status, and (c) copy number variation (CNV) in chromosomal position 6-26329011 (cytogenetic band 6p22.2) among tumor cells. CNV was inferred from gene expression with the *copykat* R package[44]. Copy number loss was also detected in this chromosomal region with targeted DNA sequencing (Supplementary DNA-Sequencing Report). **d**, Frequencies of the aneuploid and diploid variants in each intratumor maturation state in FL7. **e**, The average expression (z-scaled) and percentage of cells expressing genes with non-silent mutations detected with targeted DNA sequencing (Supplementary DNA Sequencing Report) in samples from mantle cell lymphoma (MCL), follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL) and marginal zone lymphoma (MZL). See Fig. 1 for maturation state annotations.

We also observed that genes harboring non-silent mutations differed in their expressions between intratumor maturation states. Examples included an MCL tumor with multi-hit mutated *ATM* (commonly inactivated in MCL[46,47]) showing restored expression in post-GC states, an FL tumor with multi-hit mutated *HIST1H1E* and *DTX1* (associated with transformed FL[48]) showing reduced expression only in post-DZ states, a DLBCL tumor with multi-hit mutated *PIM1* (proto-oncogene associated with ibrutinib resistance[49]) showing reduced expression only the memory state, and an MZL tumor with nonsense mutated *TNFAIP3* (NF-kB inhibitor frequently mutated in MZL[50]) showing reduced expression in post-GC states. All of these examples indicated that the mutation was acquired only in a portion of tumor cells that were either arrested or subsequently continued in maturation (Fig. 6e).

Discussion

In this study, we investigated in B-NHL how differentiation trajectories influence tumor variation and evolution. The current paradigm for B-NHL pathogenesis is that the maturation state of the tumor's cell-of-origin is a fixed determinant of its entity[11,12,43]. In contrast, we find that differentiation remains plastic in malignancy and that the differentiation trajectories guiding the nodal B-cell maturation process also drive intratumor heterogeneity.

We find that maturation states within a tumor do not always correspond with the cell-of-origin associated with its entity, revealing that intratumor heterogeneity in B-cell maturation transcends the known B-NHL entity boundaries. This deviation in maturation is bidirectional; post-GC states are observed in GC tumors (eg., FL), and, conversely, GC states are observed in post-GC tumors (eg., DLBCL non-GCB and MZL). This raises two potential roles of maturation in B-NHL pathogenesis; it may either enable additional tumor maturation states to evolve from the entity-determining cell-of-origin or drive the emergence of the predominant entity-associated tumor maturation state from a cell-of-origin earlier in the maturation process. Although it is formally possible that the B-cell maturation process acquires additional plasticity to enable shifting to earlier stages in maturation (e.g., memory or plasma to GC states) in malignancy, we consider this unlikely because many steps of the maturation process (eg., class-switching and functional differentiation) are inherently unidirectional[10]. In any case, as different B-NHL entities[14] and entity subtypes associated with different maturation states (e.g. GCB vs ABC in DLBCL)[17,51] have distinct clinical outcomes, this raises the question of whether the variation in maturation state composition among tumors is responsible for the variable disease course among patients. This is especially plausible given that the Lymph2Cx and Tally DLBCL cell-of-origin subtype classifiers, which distinguish GCB and ABC DLBCL subtypes with distinct prognoses, reveal the presence of both GCB and ABC classes within individual tumors, even in FL where a post-GC subtype has not yet been described. Thus, intratumor heterogeneity in maturation may require redefining known B-NHL entity classes to account for the various and changeable maturation states within tumors. Follow-up studies using larger cohorts will be needed to address this question. We anticipate an exciting field of research to emerge in determining its consequences and opportunities for B-NHL management.

The plasticity we observed in B-cell maturation in malignancy would enable intratumor maturation states to diverge and shift over time, thus promoting tumor evolution and expanding the selection pool for treatment resistance. Future research on the relationship between intratumor maturation states and treatment resistance could help identify resistance mechanisms and treatment stratification strategies that account for this intratumor heterogeneity and evolutionary potential. The B-cell maturation process itself may provide therapeutic opportunities, such as targeting vulnerabilities in certain maturation states or blocking tumor cells from differentiating to a resistant state. Signaling pathways and gene regulatory networks that we have shown to distinguish maturation states in tumors may serve as suitable targets, some of which already have known ligands (e.g., *XBP1*) or even approved drugs (e.g., *HDAC2*)[52]. However, such strategies should be weighed carefully against their potential for immunosuppression.

The tumor microenvironment plays an overarching role in the pathogenesis, evolution, and response of B-NHL, largely by influencing survival and growth signalings and immune infiltration or blockade[53]. The spatial organization of B-cell follicles, with germinal centers and surrounding mantle zones, is central to the normal B-cell maturation process[10,11]. We show that different intratumor maturation states occupy distinct spatial niches with unique microenvironments. Disruption of the lymph node follicular structures has long been recognized in B-NHL[54]. However, the segregation of maturation states in unique spatial niches within tumors, such as FDC and TFH co-localizing with germinal center tumor states, may explain how multiple maturation states can differentiate within a tumor. We observe a greater diversity of maturation states in FL than GCB DLBCL, which implies that retention of follicular structures better preserves the B-cell maturation process in tumors. T-cell immunotherapies, such as CAR-T cells and bi-specific T-cell engagers (BiTEs), have been gaining importance in B-NHL due to their success in relapsed or refractory DLBCL[55,56]. Varying patterns of T-cell infiltration have been observed across B-cell lymphomas[41]. We observe that immune cell infiltration patterns also vary between intratumor maturation states, including enrichment of cytotoxic T-cells among plasma tumor cells. This poses the question of whether immunotherapy may need to account for the maturation state composition of a tumor, potentially requiring combination immunotherapy to account for multiple intratumor spatial niches. The diverse microenvironments around these tumor maturation states may also encourage research on expanding the immunotherapy repertoire: leveraging not only different T-cell subtypes but also infiltrating myeloid cells and non-malignant B-cells.

The association of genetic variants with intra-tumor maturation states suggests that oncogenic mechanisms and the maturation process are intertwined, whereby certain aberrations may be more likely to lead to malignancy at different stages of the maturation process. Although inferred from single-cell transcriptomic data, the distinct gene expression patterns between maturation states associated with genetic variants detected with bulk DNA sequencing support this hypothesis. This may explain why oncogenic mechanisms have different frequencies across B-cell lymphoma entities[43]. By extending this concept to the topic of intratumor heterogeneity, one can consider that ongoing differentiation trajectories such as B-cell maturation may promote different tumor survival and proliferation mechanisms within cells of the same tumor. We expect exciting research to emerge on characterizing the association between genetic aberrations and intratumor maturation states on the single-cell level across space and time.

Similar to how sources of variation between organisms drive the evolution of species[57], sources of intratumor heterogeneity drive tumor evolution. We show that the B-cell differentiation trajectory in lymph nodes, the B-cell maturation process, is a major source of this variation in B-NHL. Intratumor heterogeneity of maturation states blurs known entity boundaries, whereby individual DLBCL and FL tumors can contain multiple previously distinct clinical subtypes. Gene regulatory networks and cellular interactions central to the B-cell maturation process are predominantly retained in malignancy, while intratumor maturation states occupy unique spatial microenvironments and may be subject to genetic variation. Seeing differentiation as a source of variation in cancer not only helps us to understand how intratumor heterogeneity arises, but also sheds light on tumor evolution trajectories and sets the stage for treatment strategies against the emergence of resistant phenotypes.

Methods

Lymph node sample processing

The University of Heidelberg's Ethics Committee approved our study (S-254/2016), and we secured informed consent from every patient beforehand. We processed and froze patient lymph node (LN) samples for later analysis, following previously described methods[41,58]. To mitigate the influence of treatment-associated effects on tumor cells and their surroundings, we excluded samples from patients who had undergone allogeneic stem cell transplantation, CAR T-cell, or bispecific antibody therapy from the CITE-Seq cohort. Furthermore, we ensured all samples were collected at least three months post the termination of the most recent treatment to maintain the same control. We provide an overview of the sample composition in Supplementary Table 1.

Single-cell 3’ RNA-seq and epitope expression profiling (CITE-Seq)

The cells were thawed, promptly washed to eliminate DMSO, and processed in groups of four to five that comprised a minimum of three distinct entities to avert entity-driven batch effects. A dead cell removal kit from Miltenyi Biotec was employed after thawing, aiming for cell viability of between 85% to 90%. Samples with less than 85% viability were not included. We then stained 5 x 10^5 live cells with a cocktail of oligonucleotide-linked antibodies (Supplementary Table 2) and left them to incubate at 4°C for 30 minutes. The cells were washed thrice with chilled washing buffer and centrifuged for five minutes at 4°C each time. Following this, cell count and viability were re-evaluated; samples falling below 85% viability were discarded. Subsequently, we prepared the bead-cell suspensions and carried out the synthesis of complementary DNA, single-cell gene expression, and the production of antibody-derived tag (ADT) libraries. For these steps, we used a Chromium single-cell v3.1 3’ kit from 10x Genomics and followed the manufacturer's guidelines (Supplementary Table 1).

Single-cell 5’ RNA-seq and B-cell receptor repertoire profiling

Apart from epitope staining, sample processing was identical to 3’ scRNA-seq. The preparation of the bead-cell suspensions, synthesis of complementary DNA and single-cell gene expression, and BCR libraries were performed using a Chromium single-cell v2 5’ and

human BCR amplification kit (both 10x Genomics) according to the manufacturer's instructions. An overview of sample libraries is provided in Supplementary Table 1.

Single-cell library sequencing and data processing

We pooled the 3' gene expression and ADT libraries in a 3:1 ratio, targeting 40,000 reads (gene expression) and 15,000 reads per cell (ADT) respectively, and sequenced them on a NextSeq 500 (Illumina). 5' gene expression libraries were sequenced on a NextSeq 2000 (Illumina), aiming for 50,000 reads per cell. BCR libraries, sequenced on a NextSeq 500 (Illumina), were aimed at achieving a minimum of 5,000 reads per cell.

Post sequencing, we utilized the Cell Ranger software's (10x Genomics, v6.1.1) *cellranger mkfastq* function for demultiplexing and aligning raw base-call files to the reference genome (hg38). For 3' gene and epitope expression libraries, we used the *cellranger count* command on the resulting FASTQ files, while we used *cellranger multi* for 5' gene expression and BCR libraries. For the BCR libraries, we used the VDJ Ensembl reference (hg38, v5.0.0) as a reference. Unless specifically stated otherwise, we adhered to default settings for all functions.

CITE-Seq data analysis

The *Seurat* R package (v4.1.0) was used to perform data quality control, filtering, and normalization (log-based normalization for RNA and centered log-ratio transformation for ADT data). Gene counts per cell, ADT counts per cell, and percentages of mitochondrial reads were computed using the built-in functions. Principal component analysis[59], Louvain clustering[60], and UMAP[61] were performed for the transcriptome (RNA) and epitope (ADT) data independently. After mapping the CD3 and CD19 epitope expression, non-B-cell transcriptomic clusters and doublets were removed. We used the IntegrateData function of the Seurat package for data integration across the different preparation batches. For multimodal clustering, multi-omic factor analysis was performed with the *MOFA2* R package[33,62] (v1.8) based on the combined transcriptome and epitope data, and the resulting latent factors (n=30) were used as principal components.

5' single-cell RNA-seq data and B-cell receptor profile analysis

Transcriptomic analysis was performed using the R package Seurat (v4.1.0) as described in the *CITE-Seq data analysis* section above. B-cell receptor (BCR) clonotypes were added to the metadata from the *cellranger multi* output.

Sorting of B-cell maturation states in reactive lymph nodes

B-cell maturation states from 5 reactive lymph nodes and 2 tonsils were identified and sorted with the below marker panels and the FACS gating strategy outlined in Extended Data Fig. 1. Maturation states and their marker panels were adapted from previous studies which defined naïve[63,64], germinal center[65], memory[63,64], and plasma[66] states with flow cytometry.

Naïve B-cells (Naïve): CD19+, CD20+, CD38 low, CD27-, IgD high
Germinal center dark zone B-cells (DZ): CD19+, CD20+, CD38+, CD184+, CD83-
Germinal center light zone B-cells (LZ): CD19+, CD20+, CD38+, CD184-, CD83+
IgM memory B-cells (Mem IgM): CD19+, CD20+, CD38 low, CD27+, IgM+
IgG memory B-cells (Mem IgG): CD19+, CD20+, CD38 low, CD27+, IgG+
Plasmablasts/plasma cells (Plasma): CD19+, CD20 low, CD38 high, CD27 high, IgD low

RNA-seq of sorted maturation states

RNA was isolated by the RNeasy Micro Kit (Qiagen, Hilden, Germany) and quantified with Bioanalyzer RNA 6000 pico assay (Agilent, Santa Clara, US). The libraries were generated with NuGENs Trio RNA-Seq System (NuGEN, Redwood City, California) for whole RNA and sequenced on an Illumina NextSeq2000 (Illumina, San Diego, US). Reads were trimmed with TrimGalore v0.6[67] and aligned with hisat2 v2.2.1[68]. The *DESeq2* R package[69] (v1.38.3) was used for differential gene expression analysis between maturation states. Default parameters were used unless otherwise specified. Gene symbols, as per the scRNA-seq datasets, were obtained from *Ensembl*[70] HGNC symbols.

Characterization of B-cell maturation states in the reactive lymph node reference

Clustering and differential expression analysis of the scRNA-seq data from B-cells in the integrated reactive lymph node samples (8 samples, 16625 cells) were performed as described in the *Seurat Guided Clustering Tutorial* [71], with the clustering resolution parameter set to 1. Clusters were assigned to known B-cell maturation states based on their differential expression (in RNA and ADT features) of established markers of B-cell maturation states from several sources in the literature[10,11,17,22–26] (Supplementary Table 3).

Classification of reactive lymph node maturation states in single-cell RNA-seq data

The RNA-seq data from the sorted maturation states was filtered by the 2000 most variable RNA features in the reactive lymph node scRNA dataset (8 samples combined) and scaled. The resulting matrix was used as input for training a logistic regression model with nested cross-validation for the classification of maturation states using the *nestedcv* package[72–75]. The resulting best-fit model was used to predict maturation states in the log-normalized and scaled scRNA-seq data from reactive lymph nodes. Predicted states were used to validate marker-based maturation state annotations in the CITE-Seq rLN reference dataset.

Mapping of maturation states in all lymph node samples

B-cell maturation states defined in the reactive lymph node reference were mapped to each tumor sample in the CITE-Seq and 5' scRNA-seq datasets using an anchor-based single-cell integration approach outlined in the *Seurat multimodal reference mapping* tutorial[30,76]. Log-normalized counts (without batch-effect correction to prevent bias introduced by sample integration) were used to find transfer anchors and project samples on the reference reductions - PCA (50 dimensions) and UMAP (2 dimensions).

Isolation of malignant B-cells

Malignant B-cells in tumor samples were identified based on immunoglobulin light chain restriction, whereby malignant (monoclonal) populations of cells are restricted to either the kappa or lambda immunoglobulin light chain and non-malignant B-cell populations (polyclonal) show mixed kappa and lambda light chain positivity[29]. As a minority of ADT counts may be present from ambient unbound antibodies during CITE-Seq library preparation, the proportion of total light chain counts of the kappa subtype (Kappa counts/(Kappa + Lambda counts)) per cell was used as a surrogate for binary positivity. Transcriptional B-cell clusters with an average kappa light chain proportion of >80% or <20% across all cells were considered malignant. Non-malignant B-cells represented a median of 6% [0%, 94%] of all B-cells in tumor samples.

Maturation state gene expression signature scoring

Maturation state gene expression signature scores were calculated by averaging the log-normalized counts for the 50 most differentially expressed genes (by fold-change) for each B-cell maturation state annotated in a published tonsil scRNA-seq dataset[17].

Inference of transcription factor activity from single-cell RNA-sequencing data

The pySCENIC[39] workflow was executed through a custom Snakemake pipeline. To infer the Gene Regulatory Network (GRN), we used the GRNBoost2 algorithm from the Arboreto[77] package with 10 perturbations. The analysis was performed on the raw scRNA-seq data. Transcription factor (TF) regulons were predicted using the human v9 motif collection from cisTarget (hg38__refseq-r80__10kb_up_and_down_tss.mc9nr.feather and hg38__refseq-r80__500bp_up_and_100bp_down_tss.mc9nr.feather databases). AUC scores per cell and GRNs were obtained for visualization and downstream analysis. For the final GRN reconstruction, only target genes occurring in more than 95% of the runs were considered. Differential expression (DE) analysis between B cell maturation stages was conducted using Seurat's FindMarkers[78] function on the RNA assay, utilizing the MAST[40] method for DE analysis from single-cell data. DE genes between conditions in all cell populations were identified (p.adj < 10e-16 & log2FC > 0.4), and p-values were adjusted for multiple comparisons using the Benjamini-Hochberg[79] correction method. To determine differentially active TFs, we utilized the output of the SCENIC GRN, which provided TF activity at a single-cell level. Differentially active TFs were detected using Fisher's exact test to assess the enrichment of maturation stage-specific DE genes among all the TF target genes extracted from the SCENIC GRN (p.adj < 0.05).

CODEX imaging and processing

Representative tumor or tumor-free lymph node areas were selected from archival FFPE tissue blocks belonging to 19 patients. This selection was made by the certified pathologists at the National Center for Tumor Diseases' Tissue Bank and the University Hospital Heidelberg's Institute of Pathology as previously described[41]. Two 4.5 mm cores per patient were incorporated into Tissue Microarrays (TMAs). TMA sections (4 µm) were affixed to Vectabond-precoated 25 x 25 mm coverslips, coated with paraffin, and stored for future staining.

Antibody conjugation, validation, and titration

We used the co-detection by indexing (CODEX) approach for multicolor immunofluorescence[21]. Antibodies utilized for CODEX experiments are summarized in Supplementary Table 2. We reduced purified, carrier-free antibodies with Tris(2-carboxyethyl)phosphine (TCEP) and conjugated them with maleimide-modified

CODEX DNA oligonucleotides, procured from TriLink Biotechnologies. A board-certified pathologist supervised the evaluation of the conjugated antibodies in singleplex stains on tonsil and/or lymphoma tissue, comparing with online databases, immunohistochemical reference stains, and published literature. We validated staining patterns in multiplex experiments with positive and negative control antibodies and titrated the appropriate dilution of each antibody starting from 1:100 to optimize the signal-to-noise ratio.

Multiplex tissue staining and fixation

We deparaffinized, and rehydrated coverslips, and subjected them to heat-induced epitope retrieval at pH9 and 97°C for 10 minutes in a Lab Vision PT module. After blocking non-specific binding with CODEX FFPE blocking solution, we stained the coverslips overnight with the full antibody panel at the dilutions shown in Supplementary Table 2. Following staining, coverslips were fixed with 1.6% paraformaldehyde, methanol, and BS3 fixative, then stored in CODEX buffer S4 until imaging.

Multicycle imaging

We attached stained coverslips to custom acrylic plates and inserted them into a Keyence BZ-X710 inverted fluorescence microscope. We selected 7x7 fields of view and an appropriate number of z-planes (10-14) to capture the best focal plane across the imaging area. Multicycle imaging was performed using a CODEX microfluidics device. Post completion of multicycle imaging, coverslips were stained with hematoxylin/eosin, and the same areas were imaged in brightfield mode.

Image processing

We processed raw TIFF images using the RAPID pipeline[80] in Matlab with the default settings. Post-processing, images were concatenated to hyperstacks. Each tissue core was visually inspected for staining quality using ImageJ/Fiji.

Cell segmentation and cell type annotation

We segmented individual nuclei based on the Hoechst stain and quantified cellular marker expression levels using a modified version of the Mask R-CNN-based CellSeg software. A threshold based on the intensity of the nuclear markers Hoechst and DRAQ5 was used to exclude non-cellular events. Cells were then submitted to Leiden-based clustering using the

scanpy Python package, and cluster annotations were assigned according to previously identified cell type marker profiles[41].

CITE-Seq to CODEX B-cell maturation state label transfer

B-cell maturation states in the CODEX data were classified sample-wise from annotations in the CITE-Seq data using shared features (n = 28) in the CITE-Seq and CODEX antibody panels (Supplementary Table 2). After selecting the shared features, CITE-Seq ADT counts and CODEX fluorescence intensities were subject to the same preprocessing steps of log-ratio normalization and scaling (z-scored) with *Seurat* v4. For each sample, a logistic regression classifier (*glmnet* package, 10-fold nested cross-validation)[74,75] was trained on the annotated CITE-seq data to classify B-cell maturation states. To prevent prediction bias toward majority classes, random sampling was performed to balance class distribution within the splits. For each sample, the resulting best-fit model (with the highest balanced accuracy on the outer folds) was used to predict B-cell maturation states in the sample's corresponding CODEX B-cell data. The median Pearson correlation coefficient between the samples' CITE-Seq and CODEX maturation state proportions was 0.91 ($p = 0.011$) (Extended Data Fig.8).

Cellular neighborhood analysis

We modified a previously described approach for neighborhood analysis[42]. For each cell, the 20 nearest neighbors were determined based on their Euclidean distance of the X and Y coordinates, thereby creating one 'window' of cells per individual cell. Next, we grouped these windows using k-means clustering according to the proportions of cell types within each window. We selected K=11 for the number of neighborhoods as we observed that higher values of k did not result in an improved biologically interpretable number of neighborhoods. Neighborhoods were annotated based on their biological function in normal lymph nodes or their enriched cell type(s)/state(s).

Cellular interaction likelihood analysis

Spatial graph representations of immediately neighboring cells were constructed based on Delaunay triangulation between centroid coordinates using the *scipy.spatial* Python package[81]. To compute pairwise association strengths between clusters, relative frequencies were computed using the following metric:

$$\frac{Nij \times Nt}{Ni \times Nj}$$

in which Nij is equal to the total number of edges between clusters i and j, Nt the total number of edges in the sample, and Ni and Nj the total degrees of clusters i and j respectively[42]. Computed association strengths were calculated separately for each disease entity, between B-cell states and other cell types.

DNA sequencing

DNA was fragmented (Covaris sonication) to 250 bp and further purified using Agentcourt AMPure XP beads (Beckman Coulter). Size-selected DNA was then ligated to adaptors during library preparation. Each library was quantified using qPCR and analyzed for quality after fragmentation and library preparation based on library yield and size on an Agilent Bioanalyzer. The sample MZL_2 failed at the library preparation stage. Finally, libraries were enriched for genes using the Sure Select XT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing and each capture pool was sequenced at 300-400x. A list of captured regions is included in Supplementary Table 4.

Pooled samples were demultiplexed using a custom demultiplexing tool. Read pairs were aligned to the hg19 reference sequence using the Burrows-Wheeler Aligner[82], and data were sorted and duplicate-marked using Picard tools (version 2.23.3)[83]. All steps were performed within the bcbio-nextgen toolkit (version 1.2.9)[84].

The minimum quality criterion was 80% of target bases having > 30x sequencing coverage. Cases with 60-79% of target bases with > 30x sequencing coverage were also included if target bases not covered were < 1%. Cases with target bases covered 30x < 60% or cases with target bases covered 30x between 60-80% and target bases not covered > 1% were excluded. This was achieved for all sequenced samples (DNA-Sequencing Report, Fig. 1). Metrics were collected using Picard tools (version 2.23.3)[83]. For detailed QC metrics see Supplementary Table 5.

Variant Analysis

Mutation analysis for single nucleotide variants (SNV) and Insertions and Deletions (InDels) was performed using MuTect2[85] (GATK v4.1.9.0)[86] and annotated by Funcotator[87] (GATK v4.1.9.0). A panel-of-normals (PON) filter was generated using samples annotated as rLN and a panel of normal from the 1000 Genomes Project[88]. Variants were included in the PON if present in two or more normal samples.

Non-silent variants (Missense_Mutation, Nonsense_Mutation, Nonstop_Mutation, Splice_Site, Translation_Start_Site) resulting from BestEffect Funcotator annotation (dataSources.v1.6) at a variant allele frequency of > 10% are kept for further investigations. Germline polymorphisms and sequencing artifacts were excluded by comparison with the panel-of-normals and with the gnomAD database[89]. Known germline polymorphisms from the Exome Sequencing Project[90] and dbSNP[91] databases were excluded. An overview of the somatic variants identified is depicted in the Supplementary DNA-Sequencing Report.

Genome-wide copy number aberrations (CNAs) were called using CNVkit (v0.9.9)[92]. Notably, this tool takes advantage of both on- and off-target sequencing reads and estimates the copy number using a pooled normal reference to compare binned read depths. Log2 change from a pool of normal control of ±0.2 was used as an indication of chromosomal gain or loss. Chromosomes X and Y are excluded from the analysis.

Inference of copy number variation from single-cell RNA-sequencing data

Copy number variants (CNVs) and ploidy were inferred from single-cell RNA-sequencing count data in each sample using the copykat R package as per the package vignette[44]. A cell filtering threshold of 5 genes per chromosome and a minimal segmentation window size of 25 genes was used. Copy number variation (Euclidean distance) was determined at a resolution of 5MB chromosomal segments, which was added as a new assay to CITE-Seq Seurat objects for each sample for visualization of copy number variants across intratumor maturation states.

Interactive data browsing

All single-cell data and tissue cores imaged in this study including the marker stainings will be available for interactive browsing on publication.

Data availability

All single-cell sequencing data will be available in the European Genome-Phenome Archive (https://ega-archive.org/) database upon publication. Highly multiplexed immunofluorescence images will be available in the BioStudies database (https://www.ebi.ac.uk/biostudies/) upon publication.

Code availability

Code scripts used for all analyses and figure generation will be made available on github.com on publication.

References

1. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
2. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613–628 (2017).
3. Roider, T. *et al.* Dissecting intratumour heterogeneity of nodal B-cell lymphomas at the transcriptional, genetic and drug-response levels. *Nat. Cell Biol.* **22**, 896–906 (2020).
4. Lin, Z. *et al.* Intratumor Heterogeneity Correlates With Reduced Immune Activity and Worse Survival in Melanoma Patients. *Front. Oncol.* **10**, 596493 (2020).
5. Turajlic, S. *et al.* Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell* **173**, 595–610.e11 (2018).
6. Yates, L. R. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.* **21**, 751–759 (2015).
7. Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).
8. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291–304.e6 (2018).
9. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
10. Morgan, D. & Tergaonkar, V. Unraveling B cell trajectories at single cell resolution. *Trends Immunol.* **43**, 210–229 (2022).
11. Seifert, M., Scholtysik, R. & Küppers, R. Origin and Pathogenesis of B Cell Lymphomas. in *Lymphoma: Methods and Protocols* (ed. Küppers, R.) 1–33 (Springer New York, 2019).
12. Koues, O. I., Oltz, E. M. & Payton, J. E. Short-Circuiting Gene Regulatory Networks: Origins of B Cell Lymphoma. *Trends Genet.* **31**, 720–731 (2015).
13. Bahler, D. W., Pindzola, J. A. & Swerdlow, S. H. Splenic marginal zone lymphomas appear to originate from different B cell types. *Am. J. Pathol.* **161**, 81–88 (2002).
14. Shankland, K. R., Armitage, J. O. & Hancock, B. W. Non-Hodgkin lymphoma. *Lancet* **380**,

848–857 (2012).

15. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
16. Shah, B. D., Martin, P. & Sotomayor, E. M. Mantle cell lymphoma: a clinically heterogeneous disease in need of tailored approaches. *Cancer Control* **19**, 227–235 (2012).
17. Holmes, A. B. *et al.* Single-cell analysis of germinal-center B cells informs on lymphoma cell of origin and outcome. *J. Exp. Med.* **217**, (2020).
18. Ayers, E. C. *et al.* Outcomes in patients with aggressive B-cell non-Hodgkin lymphoma after intensive frontline treatment failure. *Cancer* **126**, 293–303 (2020).
19. Shi, Y. *et al.* Single-cell phenotypic profiling to identify a set of immune cell protein biomarkers for relapsed and refractory diffuse large B cell lymphoma: A single-center study. *J. Leukoc. Biol.* (2022) doi:10.1002/JLB.6MA0822-720RR.
20. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
21. Black, S. *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat. Protoc.* **16**, 3802–3835 (2021).
22. Laidlaw, B. J., Duan, L., Xu, Y., Vazquez, S. E. & Cyster, J. G. The transcription factor Hhex cooperates with the corepressor Tle3 to promote memory B cell development. *Nat. Immunol.* **21**, 1082–1093 (2020).
23. Sidwell, T. & Kallies, A. Bach2 is required for B cell and T cell memory differentiation. *Nature immunology* vol. 17 744–745 (2016).
24. Cattoretti, G. *et al.* Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. *Blood* **107**, 3967–3975 (2006).
25. Goteri, G. *et al.* Comparison of germinal center markers CD10, BCL6 and human germinal center-associated lymphoma (HGAL) in follicular lymphomas. *Diagn. Pathol.* **6**, 97 (2011).
26. Delia, D. *et al.* CD1c but neither CD1a nor CD1b molecules are expressed on normal, activated, and malignant human B cells: identification of a new B-cell subset. *Blood* **72**, 241–247 (1988).
27. BioRender. Created with BioRender. com. Preprint at (2021).

28. Cox, D. R. The Regression Analysis of Binary Sequences. *J. R. Stat. Soc. Series B Stat. Methodol.* **20**, 215–242 (1958).

29. Ratech, H. & Litwin, S. Surface immunoglobulin light chain restriction in B-cell non-Hodgkin's malignant lymphomas. *Am. J. Clin. Pathol.* **91**, 583–586 (1989).

30. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).

31. Woolson, R. F. Wilcoxon Signed-Rank Test. *Wiley Encyclopedia of Clinical Trials* (2008) doi:10.1002/9780471462422.eoct979.

32. Khodabakhshi, A. H. *et al.* Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget* **3**, 1308–1319 (2012).

33. Argelaguet, R. *et al.* Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14**, e8124 (2018).

34. Scott, D. W. *et al.* Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood* **123**, 1214–1217 (2014).

35. Meyer, P. N. *et al.* Immunohistochemical methods for predicting cell of origin and survival in patients with diffuse large B-cell lymphoma treated with rituximab. *J. Clin. Oncol.* **29**, 200–207 (2011).

36. Oh, Y.-K., Jang, E., Paik, D.-J. & Youn, J. Early Growth Response-1 Plays a Non-redundant Role in the Differentiation of B Cells into Plasma Cells. *Immune Netw.* **15**, 161–166 (2015).

37. Hagman, J. & Lukin, K. Transcription factors drive B cell development. *Curr. Opin. Immunol.* **18**, 127–134 (2006).

38. Basso, K. & Dalla-Favera, R. Germinal centres and B cell lymphomagenesis. *Nat. Rev. Immunol.* **15**, 172–184 (2015).

39. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

40. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).

41. Roider, T. *et al.* Multimodal and spatially resolved profiling identifies distinct patterns of T-cell infiltration in nodal B-cell lymphoma entities. *bioRxiv* 2022.11.04.514366 (2022)

doi:10.1101/2022.11.04.514366.

42. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341–1359.e19 (2020).
43. Shaffer, A. L., 3rd, Young, R. M. & Staudt, L. M. Pathogenesis of human B cell lymphomas. *Annu. Rev. Immunol.* **30**, 565–610 (2012).
44. Gao, R. *et al.* Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* **39**, 599–608 (2021).
45. Gao, F. *et al.* Integrative genomic and transcriptomic analysis reveals genetic alterations associated with the early progression of follicular lymphoma. *Br. J. Haematol.* (2023) doi:10.1111/bjh.18974.
46. Camacho, E. *et al.* ATM gene inactivation in mantle cell lymphoma mainly occurs by truncating mutations and missense mutations involving the phosphatidylinositol-3 kinase domain and is associated with increasing numbers of chromosomal imbalances. *Blood* **99**, 238–244 (2002).
47. Yamamoto, K. *et al.* Early B-cell-specific inactivation of ATM synergizes with ectopic CyclinD1 expression to promote pre-germinal center B-cell lymphomas in mice. *Leukemia* **29**, 1414–1424 (2015).
48. González-Rincón, J. *et al.* Unraveling transformation of follicular lymphoma to diffuse large B-cell lymphoma. *PLoS One* **14**, e0212813 (2019).
49. Kuo, H.-P. *et al.* The role of PIM1 in the ibrutinib-resistant ABC subtype of diffuse large B-cell lymphoma. *Am. J. Cancer Res.* **6**, 2489–2501 (2016).
50. van den Brand, M. *et al.* Recurrent mutations in genes involved in nuclear factor-κB signalling in nodal marginal zone lymphoma-diagnostic and therapeutic implications. *Histopathology* **70**, 174–184 (2017).
51. Chapuy, B. *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat. Med.* **24**, 679–690 (2018).
52. Pharos: Illuminating the Druggable Genome. https://pharos.nih.gov/.
53. Scott, D. W. & Gascoyne, R. D. The tumour microenvironment in B cell lymphomas. *Nat. Rev. Cancer* **14**, 517–534 (2014).

54. Menon, M. P., Pittaluga, S. & Jaffe, E. S. The histological and biological spectrum of diffuse large B-cell lymphoma in the World Health Organization classification. *Cancer J.* **18**, 411–420 (2012).

55. Katz, D. A. *et al.* Open-Label, Phase 2 Study of Blinatumomab after First-Line Rituximab-Chemotherapy in Adults with Newly Diagnosed, High-Risk Diffuse Large B-Cell Lymphoma. *Blood* **134**, 4077–4077 (2019).

56. Sermer, D. *et al.* Outcomes in patients with DLBCL treated with commercial CAR T cells compared with alternate therapies. *Blood Adv* **4**, 4669–4678 (2020).

57. Darwin, C. *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*. (MVB E-Books, 2010). doi:10.4324/9780203509104/origin-species-1859-charles-darwin.

58. Roider, T., Brinkmann, B. J. & Dietrich, S. Processing human lymph node samples for single-cell assays. *STAR Protoc* **2**, 100914 (2021).

59. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**, 559–572 (1901).

60. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).

61. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).

62. Argelaguet, R. *et al.* MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 111 (2020).

63. Kibler, A. *et al.* Systematic memory B cell archiving and random display shape the human splenic marginal zone throughout life. *J. Exp. Med.* **218**, (2021).

64. Seifert, M. *et al.* Functional capacities of human IgM memory B cells in early inflammatory responses and secondary germinal center reactions. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E546–55 (2015).

65. Victora, G. D. *et al.* Identification of human germinal center light and dark zone cells and their relationship to human B-cell lymphomas. *Blood* **120**, 2240–2248 (2012).

66. Avery, D. T. *et al.* Increased expression of CD27 on activated human memory B cells correlates with their commitment to the plasma cell lineage. *J. Immunol.* **174**, 4034–4042 (2005).

67. Krueger, F. Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. *Babraham Institute* (2015).

68. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

69. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

70. Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022).

71. Seurat - Guided Clustering Tutorial. https://satijalab.org/seurat/articles/pbmc3k_tutorial.html.

72. Stone, M. An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *J. R. Stat. Soc. Series B Stat. Methodol.* **39**, 44–47 (1977).

73. Piironen, J. & Vehtari, A. Sparsity information and regularization in the horseshoe and other shrinkage priors. *EJSS* **11**, 5018–5051 (2017).

74. Friedman, J. *et al.* Package 'glmnet'. *CRAN R Repositary* (2021).

75. Lewis, M. J. *et al.* nestedcv: an R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high dimensional data. *Bioinformatics Advances* vbad048 (2023).

76. Multimodal reference mapping. https://satijalab.org/seurat/articles/multimodal_reference_mapping.html.

77. Moerman, T. *et al.* GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).

78. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

79. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).

80. Lu, G. *et al.* A real-time GPU-accelerated parallelized image processor for large-scale

multiplexed fluorescence microscopy data. *Front. Immunol.* **13**, 981825 (2022).

81. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
82. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
83. Wysoker, A., Tibbetts, K. & Fennell, T. Picard tools version 1.90. *Jhpsn*.
84. Guimera, R. V. bcbio-nextgen: Automated, distributed next-gen sequencing pipeline. *EMBnet J.* **17**, 30 (2012).
85. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv* 861054 (2019) doi:10.1101/861054.
86. *gatk: Official code repository for GATK versions 4 and up*. (Github).
87. Funcotator Information and Tutorial. *GATK* https://gatk.broadinstitute.org/hc/en-us/articles/360035889931-Funcotator-Information-and-Tutorial.
88. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
89. Karczewski, K. The Genome Aggregation Database (gnomAD). https://ibg.colorado.edu/cdrom2019/nealeB/Gnomad/boulder_190307.pdf.
90. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
91. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
92. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
93. Brown, G. R. *et al.* Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.* **43**, D36–42 (2015).

Acknowledgments

D.F. was supported by two projects from the German Federal Ministry of Education and Research (SIMONA under grant agreement number 031L0263A and SMART-CARE under grant agreement number 031L0212E), the European Research Council Synergy Grant DECODE under grant agreement No. 810296 and the PhD program of the European Molecular Biology Laboratory (EMBL). M.A.B. was supported by a Career Development award from the International Myeloma Society (IMS). T.R. was supported by a fellowship of the German Federal Ministry of Education and Research (BMBF) and a physician-scientist fellowship of the Medical Faculty of Heidelberg University. H.V. was supported by a fellowship of the German Federal Ministry of Education and Research (BMBF). N.L. was supported by the Heidelberg School of Oncology (HSO2) fellowship from the National Center for Tumor Diseases Heidelberg. O.W. is supported by an Else-Kröner Excellence Fellowship from the Else-Kröner-Fresenius Stiftung (Project-ID 2021_EKES.13) and the German Research Foundation (DFG, #WE4679/2-1). S.D. was supported by a grant from the Hairy Cell Leukemia Foundation, the Heidelberg Research Centre for Molecular Medicine (HRCMM), and an e:med junior group grant from the German Federal Ministry of Education and Research (BMBF). For the data management, we thank the Scientific Data Storage Heidelberg (SDS@hd) which is funded by the state of Baden-Wurttemberg and a DFG grant (INST 35/1314-1 FUGG). We thank Carolin Kolb (University Hospital Heidelberg), the EMBL Genomics Core Facility (Genecore), and the DKFZ Single-Cell Open Lab (scOpenLab) for their excellent technical assistance. This manuscript was edited at Life Science Editors.

Author Information

Contributions

D.F., W.H., and S.D. conceptualized the project. D.F., T.R., and M.K. performed the single-cell experiments. D.F., T.R., J.M., L.C., P.M.B., and N.L. obtained clinical documentation. D.F., T.R., B.B., J.M., and L.L. performed single-cell data preprocessing and exploratory analyses. A.K. identified B-cell maturation states with flow cytometry and sorted them for RNA sequencing. DF. and B.B. analyzed the RNA sequencing data from sorted maturation states. D.F. characterized malignancy, transcriptional phenotypes, and B-cell maturation states in the CITE-Seq and 5' BCR data. D.F. performed the maturation gene expression signature scoring. D.F. and A.M. analyzed transcription factor activity. M.A.B. performed the CODEX imaging, with preprocessing by M.A.B., H.V., and E.C. D.F., M.A.B., H.V., A.H., E.C., A.M., and F.C. characterized cell types in the CODEX data. D.F. developed and performed label transfer of B-cell maturation states from the CITE-Seq to CODEX data. A.H. performed cellular neighborhood analysis and characterized neighborhoods with D.F. E.C. and M.M.B. performed cellular interactions likelihood analysis. V.P. analyzed genetic variants in the DNA sequencing data. D.F. inferred and analyzed copy number variation in the single-cell data. W.H. and S.D. acquired most funding and resources. D.F. wrote the original draft which was initially reviewed by M.S., W.H., and S.D. All authors subsequently reviewed the paper.

Ethics Declaration

Competing interests

G.P.N. is a co-founder and stockholder of Akoya Biosciences, Inc. and inventor on patent US9909167 (On-slide staining by primer extension).

Supplementary Information

Supplementary Table 1: Overview of the CITE-Seq sample cohort and sequencing parameters.

Supplementary Table 2: Antibody panels used in the CITE-Seq and CODEX experiments, including the shared features used for CITE-Seq-CODEX data integration.

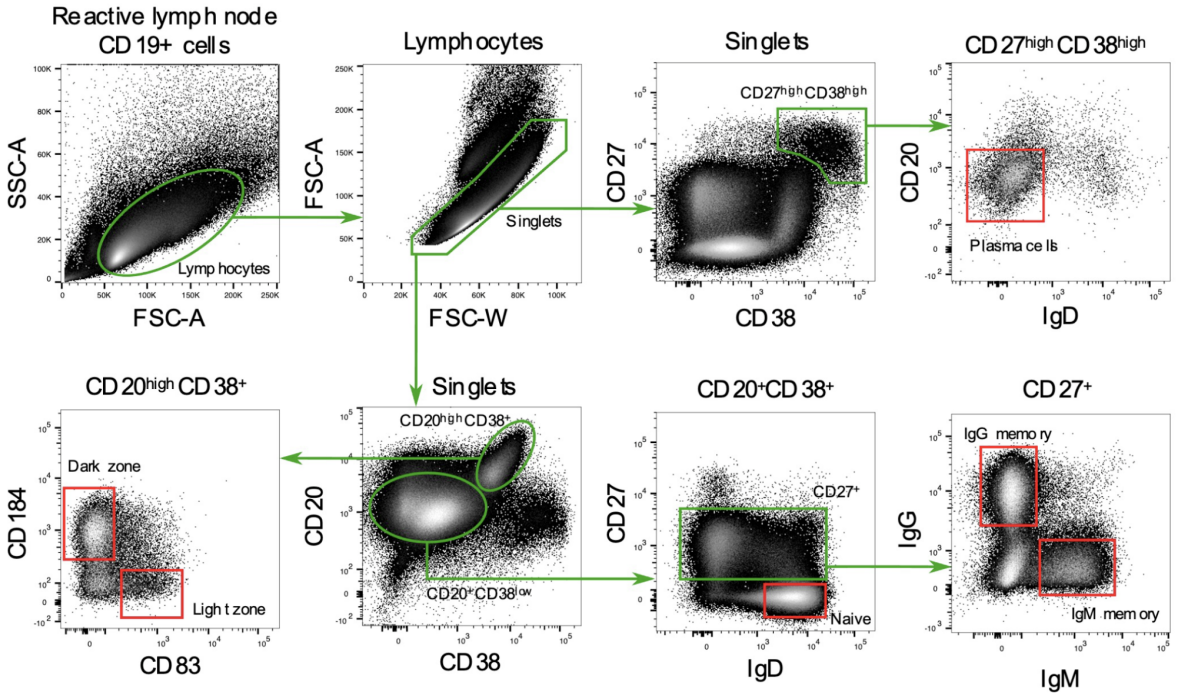
Supplementary Table 3: Referenced list of markers used to annotate maturation states and their physiological function from the NCBI gene database[93].

Supplementary Table 4: List of target genes and chromosomal positions profiled with DNA sequencing.

Supplementary Table 5: QC metrics for targeted DNA sequencing.

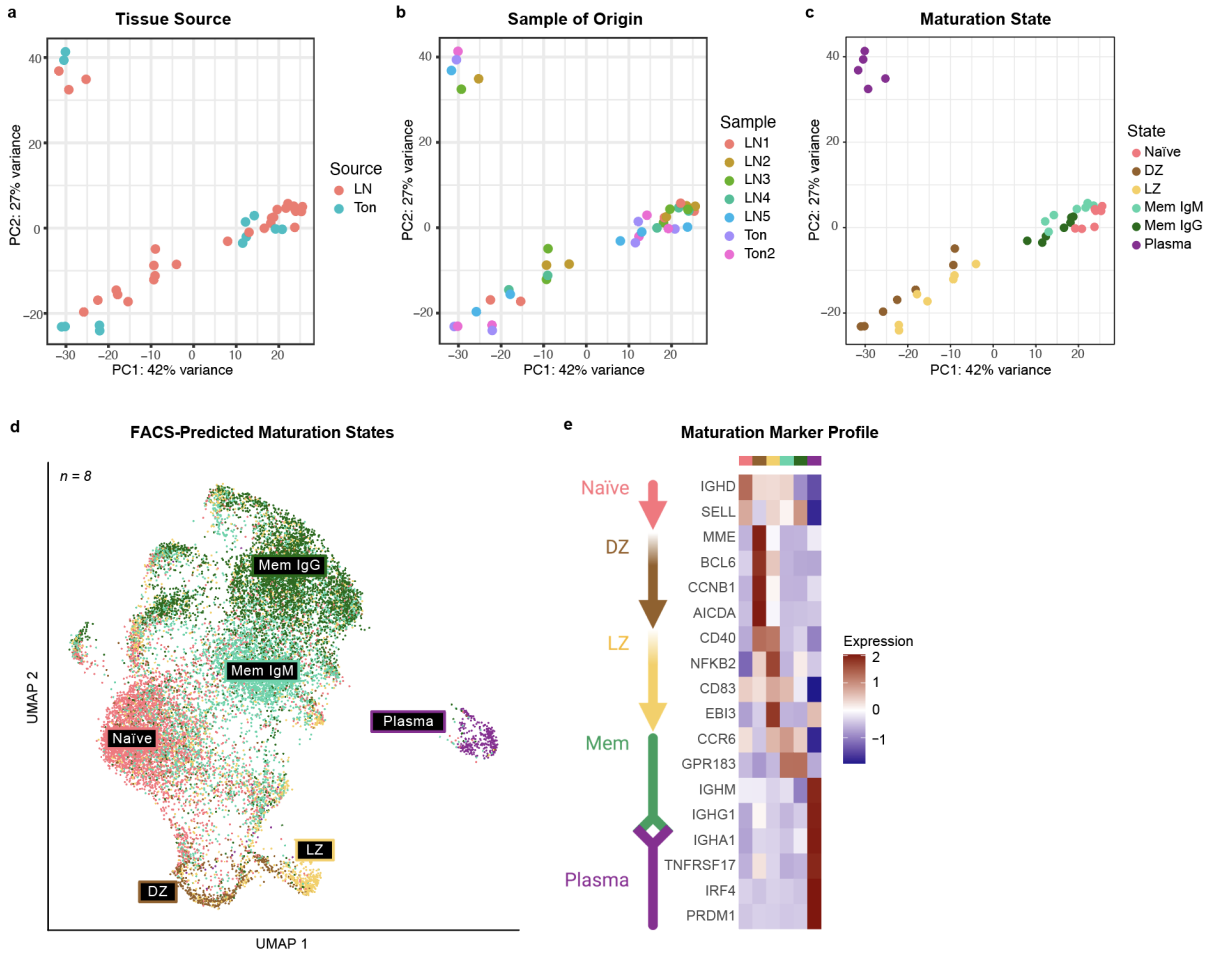
DNA Sequencing Report: Summary analysis of non-silent genetic mutations and copy number variants detected across samples with targeted DNA sequencing.

Extended Data Fig. 1: Gating strategy of B-cell maturation states from reactive lymph nodes



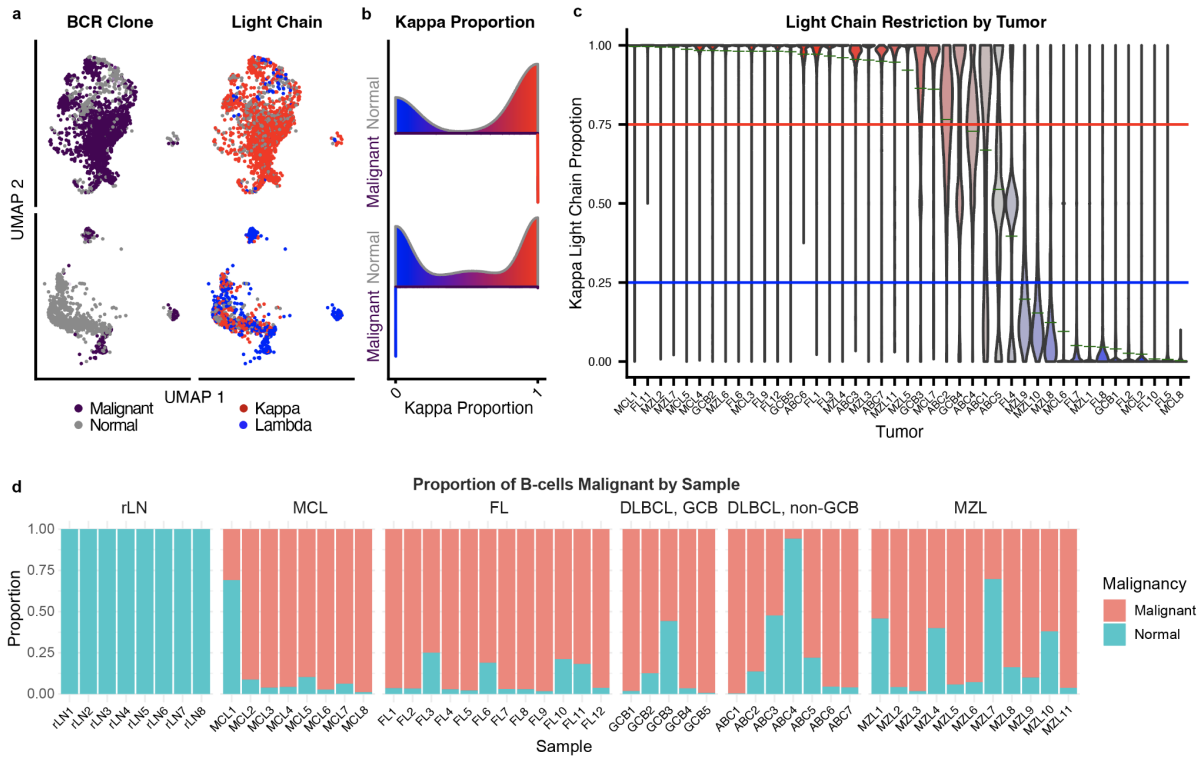
B-cell maturation state gating strategy for flow-activated cell sorting (FACS) employed on B-cells isolated from human rLN (n=5) and tonsils (n=2) using the following marker panels: Naïve B-cells (CD19+, CD20+, CD38low, CD27-, IgD high), IgM memory B-cells (CD19+, CD20+, CD38 low, CD27+, IgM+), IgG memory B-cells (CD19+, CD20+, CD38 low, CD27+, IgG+), germinal center dark zone B-cells (CD19+, CD20+, CD38+, CD184+, CD83-), germinal center light zone B-cells (CD19+, CD20+, CD38+, CD184-, CD83+) and plasmablasts/plasma cells (CD19+, CD20 low, CD38 high, CD27 high, IgD low). Maturation states and their marker panels were adapted from previous studies which defined naïve[63,64], germinal center[65], memory[63,64], and plasma[66] states with flow cytometry.

Extended Data Fig. 2: Classification from sorted B-cell maturation states



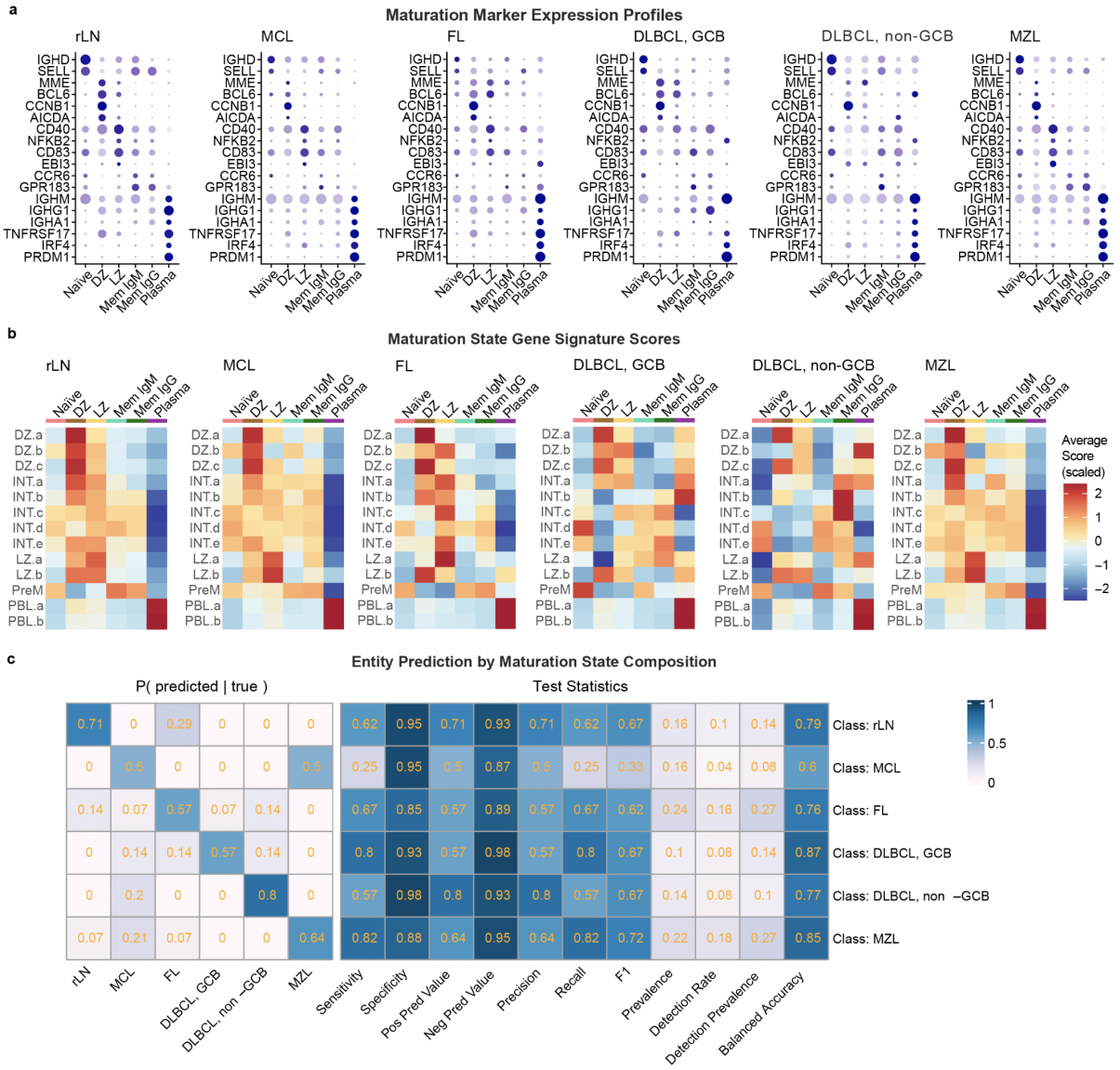
The first two principal components of RNA-seq data from FACS-sorted B-cell maturation states from rLN (n = 5) and tonsils (n = 2) colored by **a**, tissue source, **b**, the sample of origin, and **c**, maturation state. **d**, Transcriptomic UMAP of the integrated CITE-Seq B-cells data from 8 rLN labeled by maturation state predicted with logistic regression from the sorted states' RNA-seq data. e Z-scaled gene expression of a subset of B-cell maturation markers across predicted maturation states in the rLN reference. See Fig. 1 for maturation state annotations.

Extended Data Fig. 3: Isolation of malignant B-cells based on light chain restriction



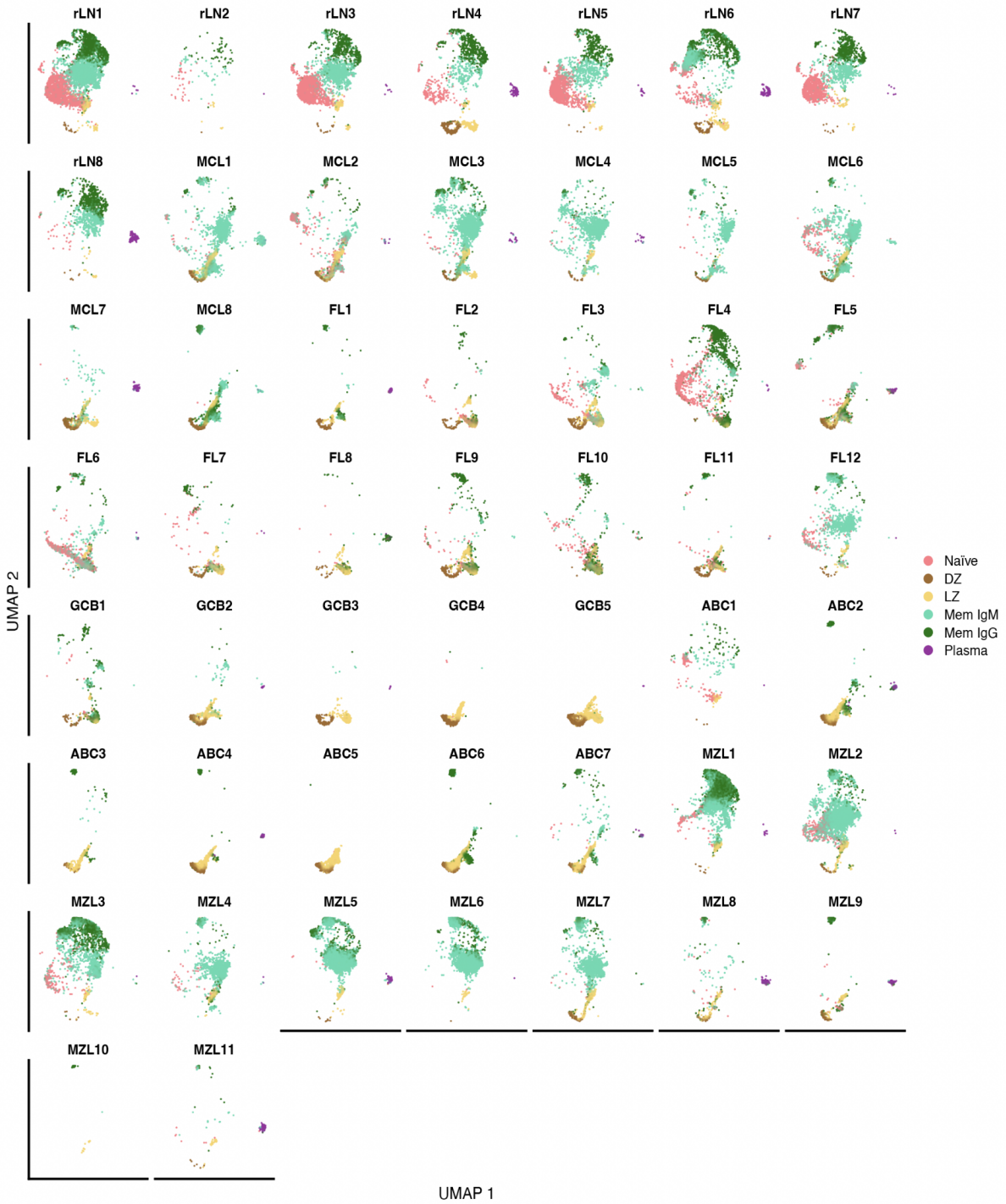
a, Reference-based UMAP labeled by malignant clone as determined by BCR profiling (left) and immunoglobulin light chain (right) for a MCL (top) and MZL (bottom) sample. **b**, Horizontal violin plot depicting the proportion of kappa light chain gene expression (x-axis) in malignant and normal B-cells in the samples shown in (a). Red = kappa-positive, blue = lambda-positive. **c**, Vertical violin plot showing the proportion of kappa light chain surface epitope detected in malignant cells isolated from all tumor samples in the CITE-Seq cohort (n = 43). Red = kappa-restricted, blue = lambda-restricted. Malignant B-cells were identified as light chain-restricted transcriptional clusters (mean kappa proportion >0.75 or <0.25). A light-chain-restricted tumor population was identified in all samples except ABC5 and FL4, which showed light-chain depletion instead. **d**, The proportion of B-cells that are malignant or non-malignant, based on light chain restriction, in each sample, faceted by entity: reactive lymph nodes (rLN), mantle cell lymphoma (MCL), follicular lymphoma (FL), germinal center and non-germinal center diffuse large B-cell lymphoma (DLBCL, GCB/non-GCB), and marginal zone lymphoma (MZL).

Extended Data Fig. 4: B-cell maturation marker expression and maturation gene signature scores by entity



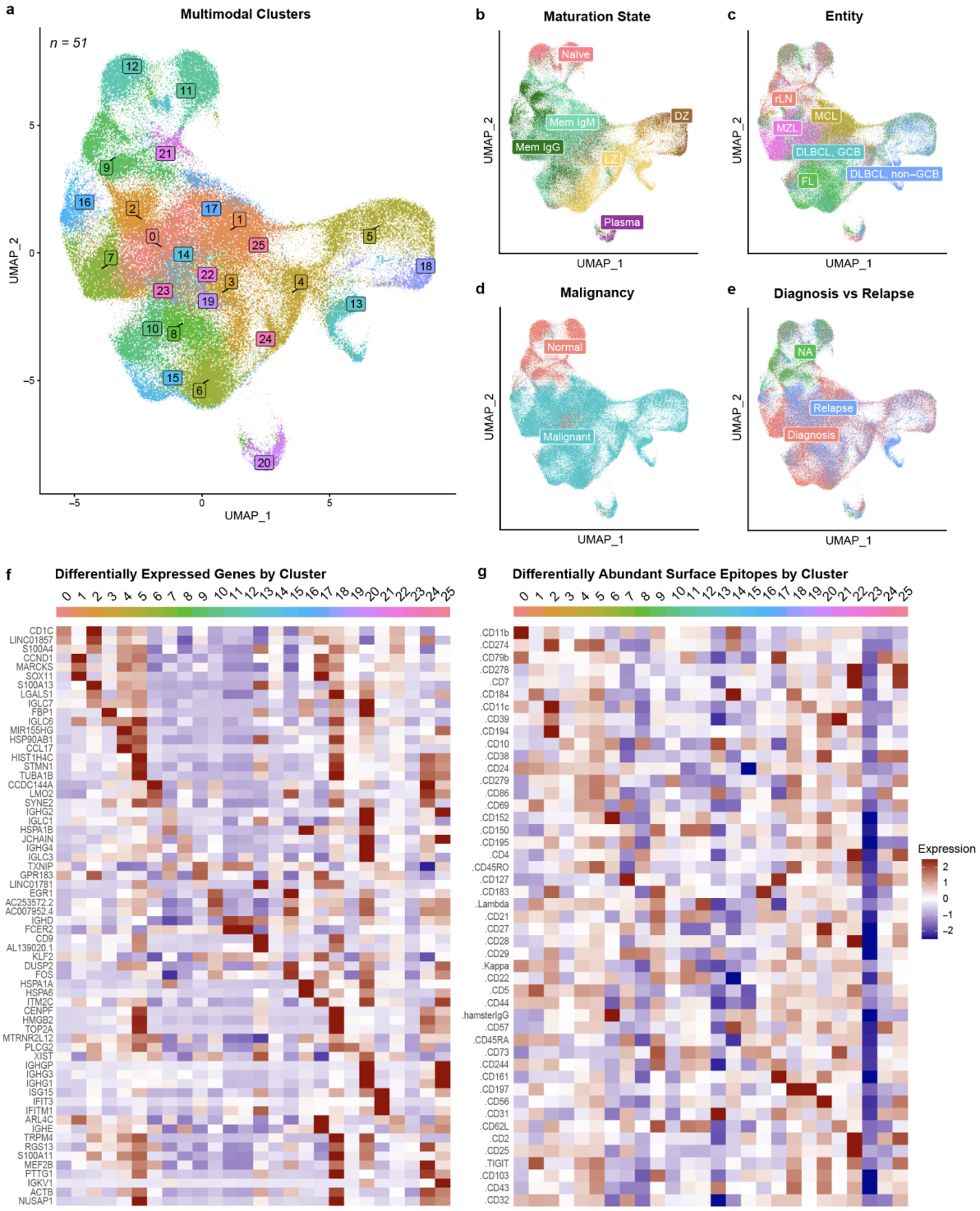
a, Dot plots showing the relative gene expression and abundance of maturation markers for each predicted maturation state in each entity. **b**, Heatmap of maturation state scores calculated from annotated maturation states in a published tonsil germinal center scRNA-seq dataset[17]. Each score shown is scaled across all scores in the dataset (mean = 0, sd = 1). **c**, Confusion matrix (left) showing the predicted (x) vs true (y) classes when predicting entity by maturation state proportions with random forest (nested cross-validation), with test statistics (right) for classification of each entity (overall accuracy 63%). Entities: reactive lymph nodes (rLN), mantle cell lymphoma (MCL), follicular lymphoma (FL), germinal center and non-germinal center diffuse large B-cell lymphoma (DLBCL, GCB/non-GCB), and marginal zone lymphoma. See Fig. 1 for maturation state annotations.

Extended Data Fig. 5: Maturation state composition of tumors



Reference-based UMAP labeled by B-cell maturation states for the CITE-Seq data from each sample (n=51). Maturation states are assigned by label transfer from the reactive lymph node reference in Fig. 1 as outlined in the Methods. Tumor cells were isolated from non-malignant B-cells based on light chain restriction of transcriptional clusters (Extended Data Fig. 3). Maturation state annotations: Naïve = Naïve B-cells, DZ = Centroblasts from the dark zone of the germinal center, LZ = Centrocytes from the light zone of the germinal center, Mem IgM = IgD+ and IgM+ memory B-cells, Mem IgG = class-switched (IgG+ or IgA+) memory B-cells, Plasma = plasma cells.

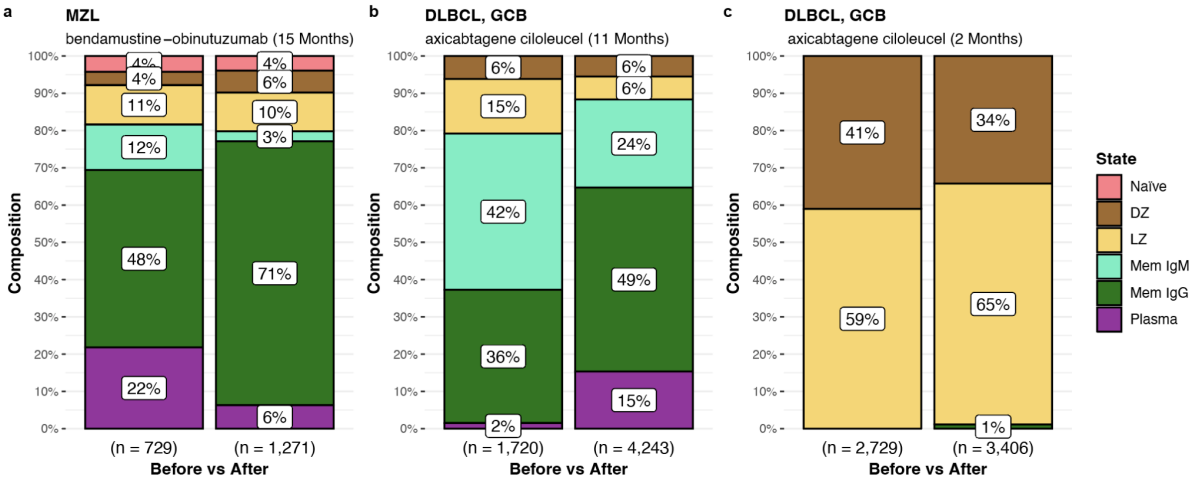
Extended Data Fig. 6: Multimodal subpopulation mapping of nodal B-cell non-Hodgkin lymphomas



UMAP visualization of the full CITE-Seq B-cells dataset (n=51) constructed with the latent factors (n = 50) from multi-omic factor analysis (MOFA)[33] based integration of the single-cell RNA and ADT (surface markers) data layers as principle components, labeled by **a**, clustering on the multimodal latent factor space, **b**, maturation states mapped from the reactive lymph node reference (Fig. 1b), **c**, entity, **d**, malignancy as determined by light chain restriction (Extended Data Fig. 2c), and **e**, samples taken at diagnosis or relapse. Z-scaled expression

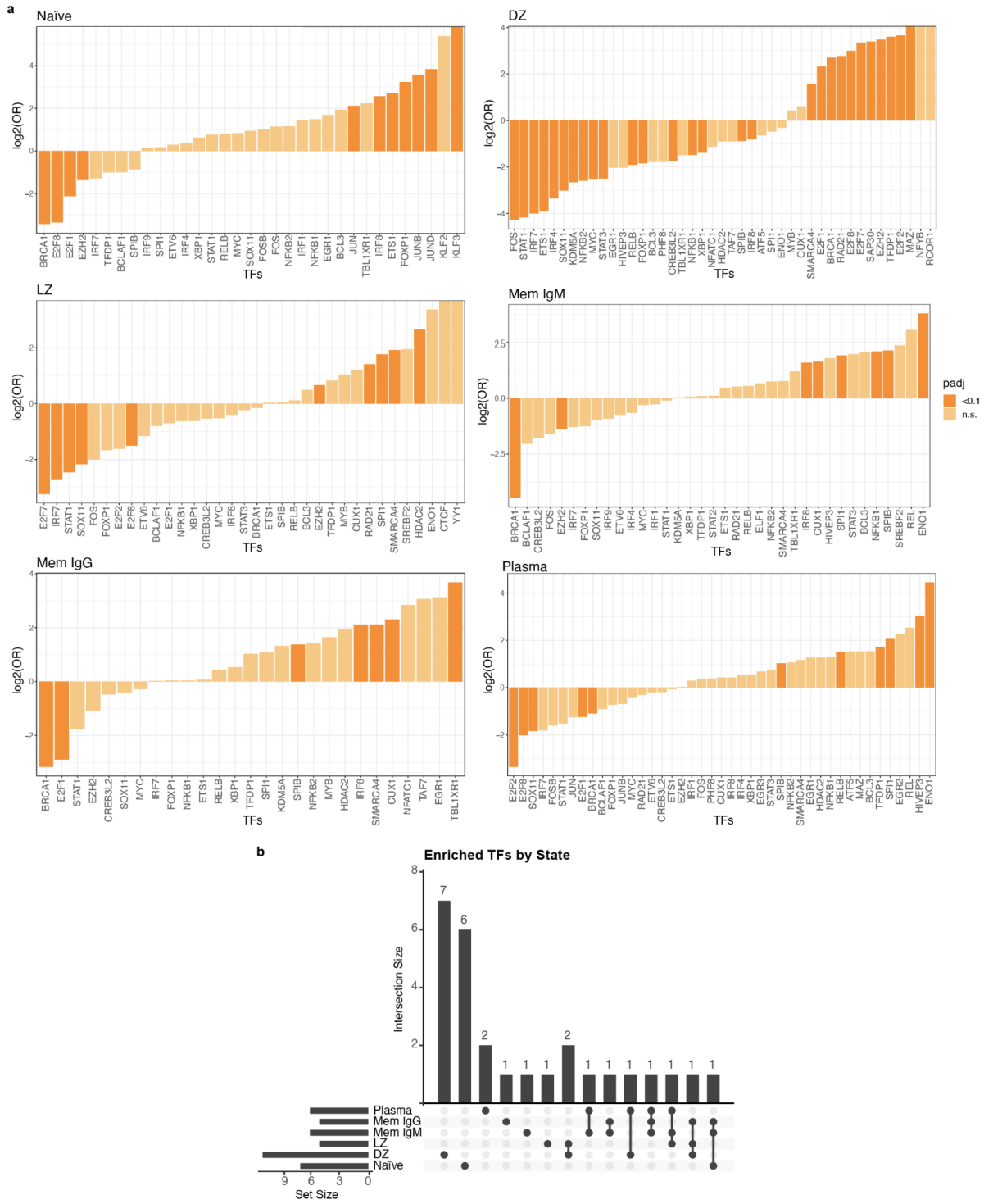
across multi-modal clusters of the 3 most differentially expressed genes (**f**) and proteins (**g**) by fold-change per cluster.

Extended Data Fig. 7: Longitudinal patterns of tumor maturation state composition



Maturation state composition of tumor cells in longitudinal samples from 3 patients: **a**, An MZL patient who relapsed 15 months following complete response to 6 cycles of obinutuzumab-bendamustine chemo-immunotherapy; **b-c**, two GCB DLBCL patients who relapsed after 11months (b) and 2 months (c) following axicabtagene ciloleucel (CAR-T cell) therapy. See Fig. 1 for maturation state annotations.

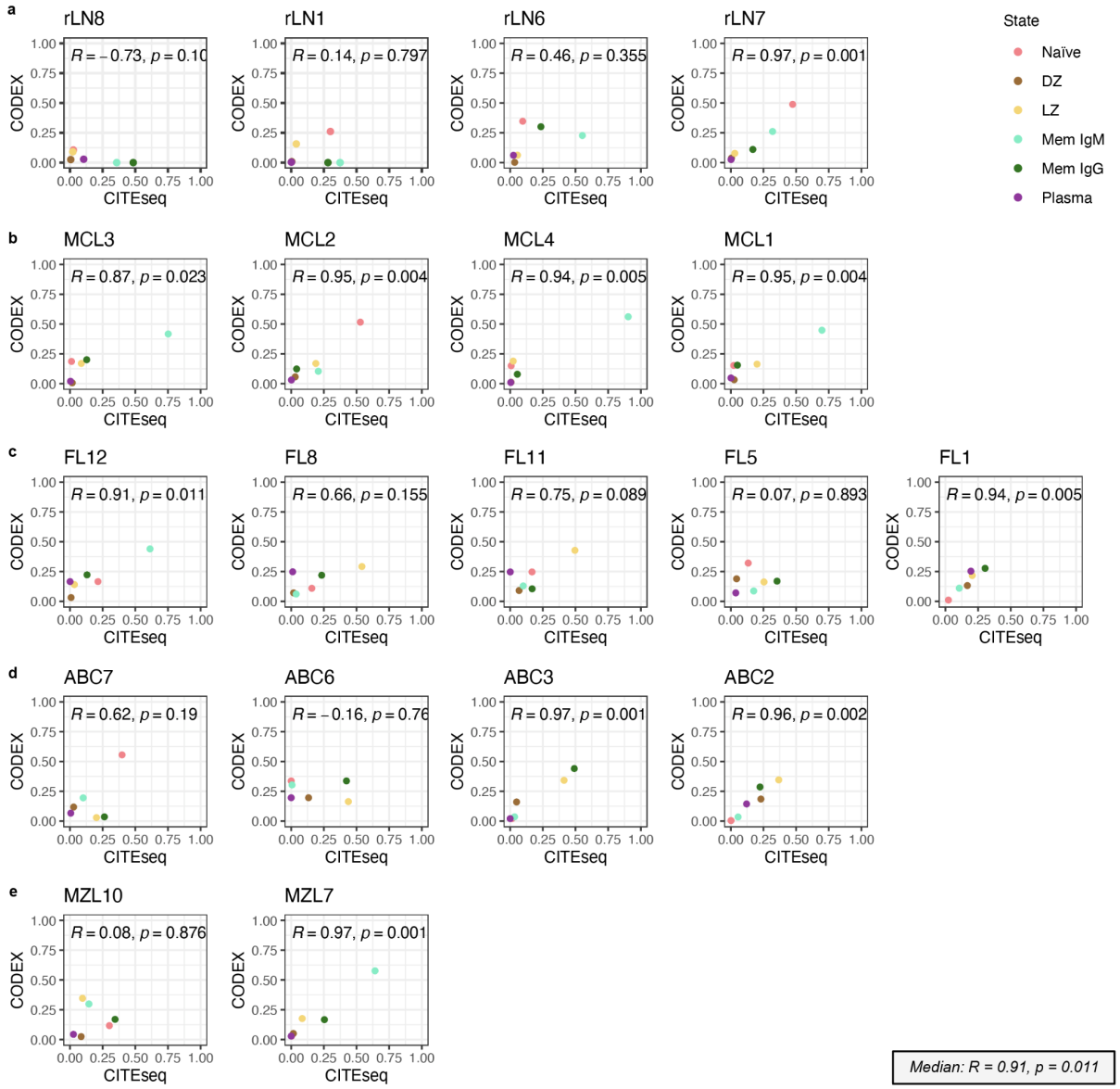
Extended Fig. 8: Differential transcription factor activity between tumor maturation states



a, Bar charts showing the log2 odds ratio for differentially active transcription factors (TFs) in the tumor cells of each B-cell maturation state inferred with the *SCENIC* python package[39] from single-cell RNA-sequencing data from the malignant cells of all tumor samples combined (n=43). Only TFs with differentially expressed target genes (log2 fold-change >0.4, p < 10e-16 as determined with the *MAST* R package[40]) are shown. Bars highlighted in darker orange represent transcription factors with significant differential activity (FDR < 0.1

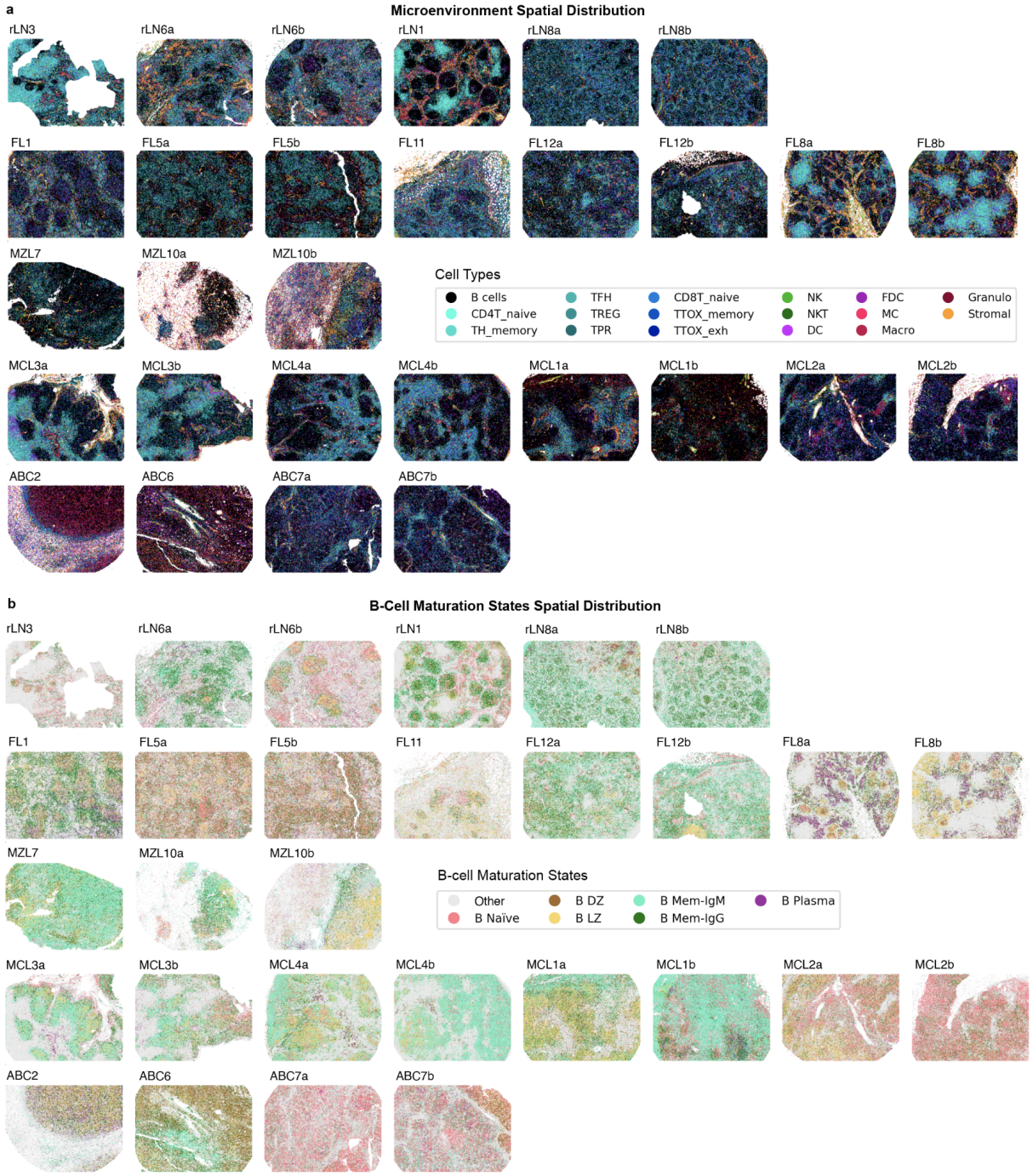
threshold). **b**, UpSet plot showing the intersections between the differentially active transcription factors (FDR <0.1) for each maturation state. See Fig. 1 for maturation state annotations.

Extended Data Fig. 9: Maturation state correlations between CITE-Seq and CODEX by sample



Scatter plots showing the Pearson correlation between B-cell maturation state proportions in the CITE-Seq and CODEX data for each **a**, reactive lymph node (rLN), **b**, mantle cell lymphoma (MCL), **c**, follicular lymphoma (FL), **d**, diffuse large B-cell lymphoma (DLBCL) and **e**, marginal zone lymphoma (MZL) sample. The median Pearson correlation coefficient across samples is shown (bottom-right). Maturation state labels were transferred from the CITE-Seq to CODEX data for each sample using logistic regression on shared protein channel/ADT features (n = 28). See Fig. 1 for maturation state annotations.

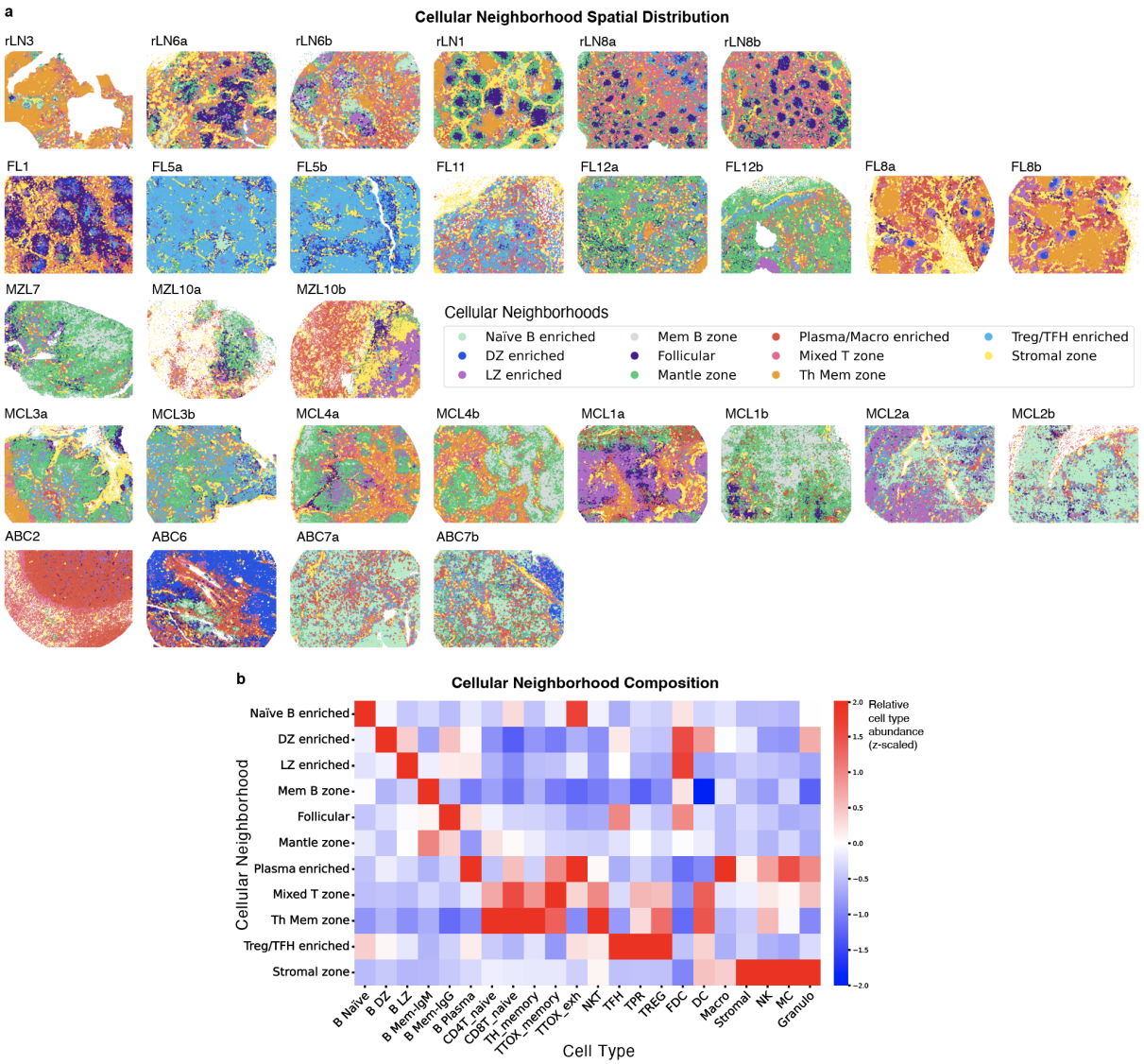
Extended Data Fig. 10: Spatial distribution of tumor microenvironments and maturation states



Spatial distribution of **a,** all cell types (with B-cells in black) and **b,** B-cell subsets in reactive lymph nodes (rLN), mantle cell lymphoma (MCL), follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL) and marginal zone lymphoma (MZL) from CODEX images on FFPE sections. B-cell states were classified samplewise using logistic regression from the CITE-Seq data using the shared features (n = 28). Marker-based annotation of all other cell types was performed on clustering of the CODEX features (n = 52, see Supplementary Table 5). B Naïve = naïve B-cells, B DZ = centroblasts from the dark zone of the germinal center, B LZ = centrocytes from the light zone of the germinal center, B Mem IgM = IgD+ and IgM+ memory

B-cells, B Mem IgG = class-switched (IgG+ or IgA+) memory B-cells, B Plasma = plasma cells, CD4T_naive = naive CD4+ T-cells, CD8T_naive = naive CD8+ T-cells, TH_memory = memory helper T-cells, TTOX_memory = memory cytotoxic T-cells, TTOX_exh = exhausted cytotoxic T-cells, NKT = natural killer T-cells, TFH = follicular helper T-cells, TPR = proliferating T-cells, TREG = regulatory T-cells, FDC = follicular dendritic cells, DC = dendritic cells, Macro = macrophages, Stromal = stromal cells, NK = natural killer cells, MC = monocytes, Granulo = granulocytes.

Extended Data Fig. 11: Spatial distribution of cellular neighborhoods



a, Spatial distribution of cellular neighborhoods (CNs) in reactive lymph nodes (rLN), mantle cell lymphoma (MCL), follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL), and marginal zone lymphoma (MZL) from CODEX images on FFPE sections. CNs (n = 11, KNN = 20) were calculated using all CODEX slides (n=29) and labeled based on their distinguishing features, ie. tumor cells' predominant maturation state (eg. DZ, Mem), B-cells' location or function in reactive lymph nodes (eg. Follicular, Mantle zone), or enriched cell type(s) (eg. Mixed T zone, T Mem zone). **b**, Relative abundance of each cell type across Cellular Neighborhoods (CNs), scaled by cell type frequency. B Naïve = naïve B-cells, B DZ = centroblasts from the dark zone of the germinal center, B LZ = centrocytes from the light zone of the germinal center, B Mem IgM = IgD+ and IgM+ memory B-cells, B Mem IgG = class-switched (IgG+ or IgA+) memory B-cells, B Plasma = plasma cells, CD4T_naive = naive CD4+ T-cells, CD8T_naive = naive CD8+ T-cells, TH_memory = memory helper T-cells, TTOX_memory = memory cytotoxic T-cells, TTOX_exh = exhausted cytotoxic T-cells, NKT = natural killer T-cells, TFH = follicular helper T-cells, TPR = proliferating T-cells, TREG = regulatory T-cells, FDC = follicular dendritic cells, DC = dendritic cells, Macro = macrophages, Stromal = stromal cells, NK = natural killer cells, MC = monocytes, Granulo = granulocytes.